CrossMark

ORIGINAL ARTICLE

Effects of Recruit Training on Police Attitudes Towards Diversity: a Randomised Controlled Trial of a Values Education Programme

Deborah Platz[1] · **Elise Sargeant**[2] · **Heather Strang**[3]

Abstract

Research Question Did a values education programme taught to Queensland police recruits change their attitudes towards police workplace diversity and equality, relative to recruits in the same cohorts who did not receive the programme?

Data A survey designed to measure attitudes towards workplace diversity and related issues was administered three times to 260 police recruits, who were randomly assigned to receive a values education programme or not over the 25-week initial police recruit course. The surveys were conducted in week two of the course, at the conclusion of the values education programme and six weeks after the programme concluded.

Methods Three separate cohorts were split by batch random assignment into experimental and controls, for 132 experimental recruits and 128 controls. Using a variety of validated scales and items, the attitudes of the two groups were compared at all three survey waves and in comparative longitudinal trends.

Findings While the values education programme did not improve experimental group recruit attitudes towards diversity in the workplace over time, it protected that group from a clear decline in support for diversity associated with the standard recruit training experience. Because the design was a randomized controlled trial (RCT), the study clearly revealed that the benefit of the programme was as a successful buffer against what happened to reduce diversity support among the other recruits.

Conclusions The findings show that in at least one police recruit experience, there is a clear shift away from support for diversity by race and gender in the police workplace

✉ Heather Strang
hs404@cam.ac.uk

1   Australian Federal Police, Canberra, Australia

2   School of Criminology and Criminal Justice, Griffith University, Brisbane, Australia

3   Institute of Criminology, University of Cambridge, Cambridge, UK

🖄 Springer

in the course of initial training. Fortunately, the results also provide at least one possible preventative measure for that problem, in the form of a values education programme similar to one used widely in many countries.

Introduction: Values Education in Policing

Public debates about police behaviour often centre on issues of bias and diversity. Much of the debate is adversarial, in the sense that the argument hinges on a verdict about whether or not police support equality of race, gender and religion. Much less of the debate is constructive, in the sense of considering ways that police can enhance their support for diversity and equality in diverse societies. The objective of this study is to contribute to the constructive literature on policing and diversity.

The study set out to evaluate the impact of a globally available values education programme, using a rigorous research design: a randomised, controlled trial. What it discovered was not only a tested benefit of the specific programme, but—in the absence of this special programme—a clear reduction in recruit support for diversity from entering to leaving the police academy. The major benefit of the values education programme was to hold recruit attitudes at the level of support for diversity they enrolled with, while recruits receiving only the standard curriculum were tracked to have significant reductions in their support for diversity.

An introduction to these findings requires a basic assessment of what was known about these issues prior to the start of the experiment.

Voice for Values (V4V)

The V4V programme (Polanin et al. 2012) teaches participants to recognise and intervene in poor workplace behaviours. Having piloted the programme on five occasions with strong positive assessments, the Queensland Police decided to conduct a randomised controlled trial (RCT) evaluating the effectiveness of the programme. There were three main drivers of this trial: firstly, to reduce poor behaviours within the Queensland Police Service (QPS); secondly, to ensure the programme worked in the way envisaged; and thirdly, due to the expense of the programme, to examine whether it was cost-effective.

The QPS academy trains approximately 800 recruits each year. With literature suggesting that recruits' values decline during their training (Ford 2003; Chan et al. 2003) and the fact that this training venue is ideal for moral training (Sherman 1980), recruit training was thought to be the best location to conduct an RCT. Over three intakes in 2015, 260 recruits entered into the RCT. Randomised into experimental and control groups, the experimental group participated in the V4V programme. The second author of this article developed a survey to test the logic model designed for this training.

The model worked on the theory that participants in the V4V programme needed to have enhanced recognition of poor behaviours, expressing racism or sexism, in order to

encourage them to intervene in such incidents. Recruits completed Dr. Sargeant's survey of their attitudes towards diversity and related issues at three points in time: (1) commencement of the recruit training programme, (2) immediately after the V4V programme and (3) six weeks after the intervention. The survey analysed the recruits' ability to recognise prejudicial, racist and sexist behaviour; acceptance of equality and diversity; enhanced empathy and stated willingness to intervene in racist and sexist incidents.

The Voice 4 Values (V4V) programme was designed especially for the QPS to train recruits to recognise and understand harms in the workplace and foster values that encourage them to intervene in workplace harassment incidents.

In an early survey (Circelli 1998) of 900 QPS officers, 92% of female police officers and 67% of unsworn female members reported experiences within the preceding two years of at least one harassing behaviour. More recently, the Victorian Equal Opportunities and Human Rights Commission (VEOHRC) (2015) investigated harassment within the Victoria Police. Over 5000 members took part in the research, which found that 40% of women and 7% of men had experienced sexual harassment (VEOHRC 2015). The research found that sexual harassment was normalised and minimised and that sexism and poor behaviours were widespread across the organisation (VEOHRC 2015).

Research shows that when people apply to become police officers they have a willingness to help people (Cumming et al. 1965; Ford 2003) and demonstrate high ideals and values (McNamara 1967; Ford 2003). However, during their time at the academy (Ford 2003), followed by their first year of field training (Sherman 1980) and their subsequent years as a sworn officer (White et al. 2008), such values and attitudes can erode after being exposed to negative aspects of police culture.

Police often witness poor behaviours, but many incidents are never reported or intervened in (Crime and Misconduct Commission 2010). Polanin et al. (2012) suggest one reason for this is that employees need to recognise poor behaviours in the workplace in order to prevent them. Training, often viewed as the panacea to all problems, is one way to ensure recognition and intervention of poor behaviours, but training aimed at enhancing one's ethics and values is difficult (More and Wegener 1992; Piper et al.1993).

V4V is intended to provide recruits with the knowledge of harms in the workplace and encourage them to have a voice to intervene. The relationship between the programme and its aims combined with the recruit values and their attitudes and perceptions of behaviours in workplace provide the basic logic model (Fig. 1). The desired outcomes of the logic model include enhanced recognition of poor behaviours and a stated willingness to intervene in them.

The V4V logic model outlines two active components that shape a recruit's attitudes and perceptions towards recognition and acceptance or non-acceptance of poor behaviours in the workplace such as racist and sexist behaviours, prejudice, and lack of equality and diversity, along with empathy and stated willingness to intervene. These active components are required for effective prevention of these behaviours, because people must have both the knowledge and a willingness to do or say something (Polanin et al. 2012). The first active component in the logic model is knowledge of workplace harms. The second active component considered to influence these attitudes and perceptions is encouragement to have a willingness to intervene.
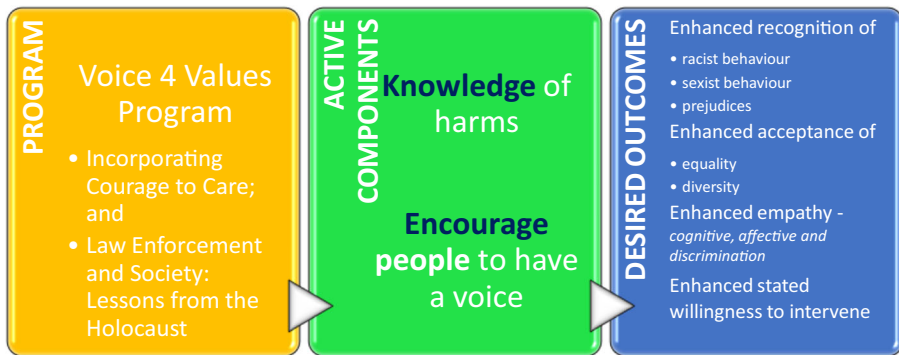
Fig. 1 Voice4Values logic model

Until the V4V programme, the QPS offered no formal training specifically focused on racist and sexist attitudes, prejudice, equality, diversity or empathy. V4V was designed to fill this gap in recruit training. V4V explicitly seeks to increase knowledge of workplace harms and encourage participants to have a voice to stand up to workplace harms. It was developed for the QPS from two other related programmes—"Courage to Care" (a school-based programme conducted by the Courage to Care organisation) and "Law Enforcement and Society: Lessons from the Holocaust" (a programme conducted by the Federal Bureau of Intelligence (FBI) and the Anti-Defamation League; see Platz 2015).

Teaming with Courage to Care, and utilising a similar format to the US programme, the QPS developed V4V. This programme seeks to instil in police the ability to recognise prejudice and racist and sexist behaviour in the workplace, along with the importance of diversity, equality and empathy. It then builds upon this knowledge to encourage individuals to make a difference and to intervene when they are subjected to or a witness to poor behaviours. V4V educates people of the dangers of prejudice and discrimination through stories of survival of the Holocaust from victims, perpetrators and bystanders.

Differing from the US programme, the V4V programme includes the additional constructs of sexist behaviour and empathy to address negative components of police culture. In V4V, police recruits view a DVD depicting scenarios of a variety of poor behaviours in a policing workplace. These scenarios were compiled from a list of regular complaints made to the Ethical Standards Command of the QPS. The DVD portrays three inappropriate workplace scenarios. They relate to racism—the false belief that certain racial groups are better or worse than others (Kleg 1993); prejudice—the irrational attitudes, beliefs and opinions that the members of one group have for another (Kleg 1993); and discrimination—the unjust treatment of people, especially on the grounds of race, gender, age, sex or religion (Ellis and Watson 2012). The incidents contain examples of sexual discrimination, bias, and racist and prejudicial language. The scenarios, also posed as vignettes in the survey, provided realistic events for recruits to consider.

Following the DVD, the focus shifts to an examination of law enforcement during the Holocaust. This component of the program begins with an historian presenting a brief audio-visual history of World War 2. The presentation focuses on the way in which law enforcement authorities were encouraged to comply with orders from above and draws on real life situations and people who intervened in immoral situations during this time. Through this presentation participants consider how police could commit atrocities and the likelihood of this happening today.

Participants then hear from a death camp survivor who describes his experiences with law enforcement, including stories of how officers aided and hindered his family's survival. Through group facilitation, the recruits are guided through these real-life examples to consider how they would react in similar circumstances (Harris 2004). This component of the V4V involves the theory of role-playing: through understanding others, role-playing can produce changes in people's attitudes and opinions (Harris 2004). A professional facilitator concludes the training with an interactive session.

From an economic perspective, V4V is a costly programme to deliver. It relies on an interstate organisation travelling, with an historian, a survivor and a specialist facilitator. It also necessitated the production of a DVD. The estimated cost of this programme is $480 per person and delivery to the nearly 16,000 members of the QPS could cost $7.6 million. Conducting an RCT is viewed as the most reliable method for determining whether the treatment is effective (Weisburd 2010).

Research Question

While a number of outcomes have been analysed from the test of V4V, this article focuses specifically on the effects of VFV on recruit attitudes towards diversity and equality in the workplace. Our specific research question in this article is *whether the values education program taught to Queensland police recruits changed their attitudes towards police workplace diversity and equality, relative to recruits in the same cohorts who did not receive the program.*

This article reports on the RCT tests of the following hypotheses:

- H1 That recruits receiving the V4V programme are more likely to recognise prejudice and racist and sexist behaviour in the workplace than those recruits not receiving the V4V training.
- H2 That recruits receiving the V4V programme are more likely to have an enhanced acceptance of equality and diversity in the workplace than those recruits not receiving the V4V programme.

Defining Attitudes Supporting Equality and Diversity Attitudes supporting diversity are indicated by recognition and valuing of differences among individuals and placing a positive value on the difference they bring. Attitudes supporting equality are measured by support for treating everyone fairly and giving them equal access to

opportunities (O'Brien 2011). Dimensions of diversity include gender, age, language, ethnicity, sexual orientation, religious beliefs, cultural background and family responsibilities (Cox and Blake 1991).

Can Training Impact These Attitudes? Research into police culture is prolific. It considers measures of racism, prejudice, harassment and other ethical dilemmas. But research evaluations of *intervention* programmes in police agencies are hard to find. This differs from the field of education where intervention programmes, especially those concerning prejudice, racism and bullying, have been evaluated (Polanin et al. 2012). A meta-analysis of 12 school-based bullying prevention programmes involving 12,874 students across Europe and the US, and the effects the programmes had on bystander intervention behaviour, yielded evidence to suggest that the programmes increased people's willingness to intervene in experimental groups as compared to the control groups, despite differences in locations, ages, treatments and cultures (Polanin et al. 2012). The most effective programmes were those using the active witness model (Polanin et al. 2012). Yet in the domain of policing, an RCT assessing the impact of specific training that seeks to enhance workplace attitudes and values could not be located.

Data

The Research Site

The QPS academy trains approximately 800 recruits each year. Recruits enter into the academy at various times throughout the year. Each "intake" consists of two to six squads, with approximately 24 recruits in each squad. Recruits undertake a 25-week training course designed to develop competent, ethical, efficient and effective police officers (*www.policerecruit.qld.gov.au*). After successfully completing their training, recruits are inducted as constables and begin a field-training programme.

Experimental Design and Randomisation

In May, July and August 2015, 260 police recruits, randomised by intake group, entered into training. The May intake had 47 recruits in two experimental squads and 45 recruits in two control squads. The July intake had 43 recruits in two experimental squads and 41 recruits in two control squads. The August intake had 42 recruits in two experimental squads and 42 recruits in two control squads. Overall, there were six experimental squads ($n = 132$) and six control squads ($n = 128$). All 260 recruits completed a pre-intervention survey, which serves as the baseline data. The experimental group received the intervention V4V programme after being in the academy for approximately two weeks. Immediately following the intervention, both the control and the experimental groups participated in a first post-intervention survey. The recruits were surveyed again six weeks post-intervention.

Survey Method

This RCT utilised a survey (see Platz 2016 appendix 1) designed by the second author, who at the time was a Lecturer in Criminology at the Institute for Social Science Research, University of Queensland (UQ). The survey and the entire V4V experiment received approval from both the UQ and QPS Ethics Committees. The survey was uploaded to an online survey software product, Qualtrics. To ensure anonymity, recruits designed their own unique identification number. It is not possible to trace the survey results to an individual recruit; however, the unique identification numbers allow tracking of results across the three surveys.

Survey Constructs

The survey was designed to measure a number of possible outcomes of the intervention, including the impact of the constructs of racism, sexism, empathy, tolerance for diversity, prejudice and discrimination, empathy, organisational legitimacy, and a stated willingness to intervene in sexist or racist behaviour in the workplace. Each construct was measured with a series of questions developed utilising a variety of scales and instruments. Most responses used a 5-point scale from "strongly disagree" to "strongly agree".

Racist Behaviour Being able to recognise racism is important, and this type of behaviour was a targeted scenario in the V4V programme. The V4V survey sought to measure the capacity of recruits to recognise racist behaviour as unacceptable by asking questions adapted from survey questions developed and used by Pennay and Paradies (2011). These arose from the Framework to Reduce Race-based Discrimination and Support Diversity in Victoria (VicHealth 2009) and a review paper (Nelson et al. 2010). Questions were also based on the Confronting Prejudiced Responses model developed by Ashburn-Nardo et al. (2008). The questions were also informed by focus groups, a series of cognitive interviews and formal pilot testing. Pennay and Paradies (2011) found participants were willing to answer the direct questions concerning racism during pilot stages and during the research.

The questions asked in the V4V survey were the following:

How acceptable would it be for an employee of the Queensland Police Service to engage in one of these behaviours?

– Using racist language to describe a work colleague;
– Using racist language to insult or abuse a work colleague;
– Telling a racist joke.

The response options for each of these three scenarios were never acceptable, rarely acceptable, sometimes acceptable and always acceptable. For each of the scenarios, participants scored one when they indicated that such behaviour was never acceptable, and zero for any other response. These three scores were then added together to form a composite scale, with higher values (up to a maximum of a score of 3) indicating an increased willingness to recognise racism in the workplace.

Sexist Behaviour The survey included questions about sexist behaviour towards women, discrimination, the unfair treatment of women and violence towards women. Pennay and Powell (2012) found that participants were prepared to respond to the questions in the pilot and research stages. The V4V survey sought to measure the capacity of recruits to recognise sexist behaviour as unacceptable by asking questions adapted from survey questions developed and used by Pennay and Paradies (2011). The questions asked in the V4V survey were the following:

How acceptable would it be for an employee of the Queensland Police Service to engage in one of these behaviours?

–   Using sexist language to insult or abuse a work colleague;
–   Using sexist language to describe a work colleague;
–   Telling a sexist joke.

The response options for each of these scenarios were: never acceptable, rarely acceptable, sometimes acceptable and always acceptable. For each scenario, participants scored one when they indicated that such behaviour was never acceptable, and zero for any other response. These three scores were then added together to form a composite scale, with higher values (up to a maximum of a score of 3) indicating an increased recognition of sexism in the workplace.

Prejudice Some of the most obvious examples of prejudicial behaviours are those based on gender, ethnicity, disability, sexual preference and age (McLeod 2008). These behaviours are the basis of scenarios in the V4V programme and tested with the survey. To create a measure of prejudice, the Social Dominance Orientation (SDO) measure developed by Pratto et al. (1994) was adapted. The SDO seeks to measure "the extent to which one desires that one's in-group dominate and be superior to outgroups" (Pratto et al. 1994, p.742). The full SDO measure has a Cronbach's alpha score of 0.83. The questions asked in the V4V survey were as follows:

Beside each object or statement, please select the response which represents the degree of your positive or negative feeling:

–   Some groups of people are simply not the equals of others;
–   Some people are just more worthy than others;
–   Some people are just more deserving than others;
–   Some people are just inferior to others;
–   To get ahead in life, it is sometimes necessary to step on others.

The response options were "very negative" (scored 7) to "very positive" (scored 1). These scenarios had a Cronbach's alpha score of 0.84. For each of the scenarios, the participant's score was recorded and then the composite measure was created by adding up all scores and dividing by 5. A higher score for the composite measure represents a less prejudiced attitude.

Equality To create a measure of equality, the SDO measure developed by Pratto et al. (1994) was again adapted. The full SDO measure has a Cronbach's alpha score of 0.83. The questions asked in the V4V survey were as follows: Beside each object or

statement, please select the response which represents the degree of your positive or negative feeling.

- If people were treated more equally, we would have fewer problems in this country;
- In an ideal world, all nations would be equal;
- We should try to treat one another as equals as much as possible (all humans should be treated equally);
- It is important that we treat other countries as equals.

The response options were "very negative" (scored 1) to "very positive" (scored 7). These scenarios had a Cronbach's alpha score of 0.81. For each of the scenarios, the participant's score was recorded and then the composite measure was created by adding up all scores and dividing by 4. A higher score for the composite measure represents greater belief in equality.

Tolerance of Diversity Nakui et al. (2011) developed a comprehensive measure, Attitudes Towards Diverse Workgroup Scale (ATDWS) to measure an individual's experience at the work group level. The overall Cronbach's alpha score for the full ATDWS is 0.84. The questions asked in the V4V survey were as follows:

Thinking about your experience in your workplace group, please indicate your level of agreement with the following statements:

- I find interacting with people from different backgrounds very stimulating;
- The experience of working with diverse group members will prepare me to be a more effective employee in an organisation;
- Diverse groups can provide useful feedback on one's ideas.

The response options were "strongly disagree" (scored 1) to "strongly agree" (scored 5). These scenarios had a Cronbach's alpha score of 0.79. For each of the scenarios, the participants score was recorded and then the mean score was created by adding up the three scores and dividing by 3. A higher score for the mean measure represents greater belief in the value of diversity.

Analytic Methods

The data from the experimental and control groups were analysed using a significance test to determine the impact of the V4V intervention on recruit ability to recognise racist and sexist behaviour, prejudice and diversity. Independent sample (unpaired) t tests, also referred to as two-tailed t tests, were used to determine if there was a relationship between variables in either direction (Bachman and Schutt 2014). This test analyses the difference in the two independent samples, the experimental and the control groups. This test was used because it analyses the responses of participants in both directions of the normal distribution and determines the statistical significance or otherwise of the impact of V4V. The level of significance applied in this research is $p = 0.05$, which

means that there is a 5% probability that the outcome is a chance occurrence. This is a widely accepted convention in the social sciences for nearly 100 years (Bross 1971). In this study, it is recognised that about 1 in 20 of the tests could return a significant result when there is, in fact, no effect.

The Cohen's *d* equation determined the *effect size* of any differences, or "how big" the difference was (Ariel and Sherman 2014). This idea is different, and perhaps more important, than the question of "how sure" we can be that this finding is not a fluke–which is the key idea underlying statistical significance. The guidelines of small (0.2), medium (0.5) and large (0.8) effect sizes suggested by Cohen (1977) were used in interpreting the effect of the V4V intervention. Considering effect size is essential when considering the costs versus benefits of implementing the V4V programme. While significance testing may show that V4V is effective in enhancing recruits' values and stated willingness to intervene, the practical benefit of the change may be negligible when compared to considerations such as the financial cost implementing the programme, or the loss of other training that this programme may replace.

Response Rates for Survey

The response rates for all stages of the survey are shown in Table 1.

Over the three surveys, the response rate was 82% for the experimental group and 89% for the control group. Completion of the survey was voluntary and could not be audited.

Findings

This section presents the survey results at three points: baseline, immediately post-intervention (follow-up 1) and then six weeks post-intervention (follow-up 2). For all constructs measured in this chapter, a higher score indicates a greater awareness of poor behaviours.

Baseline Results

At baseline, the two groups (experimental and control) were equivalent on all test measures—just as expected due to the random allocation of recruits the groups. Using a two-tailed *t* test for all constructs examined in this RCT, Table 2 presents results comparing the experimental and control groups at baseline.

There were no statistically significant differences ($p < 0.05$) between the mean scores for the control and experimental groups for any of the constructs at baseline.

Table 1 Response rates for experimental and control groups over time

Group	Baseline (%)	Follow-up 1 (%)	Follow-up 2 (%)	Overall (%)
Experimental group	84	82	79	82
Control group	99	98	70	89

Table 2 Data and statistical analysis of constructs in survey at baseline

Constructs	Control mean (score)	N	Experiment mean (score)	n	t	Df	p	Cohen's d
Recognise racist behaviour	2.33	127	2.35	111	− 0.19	236	0.85	0.024
Recognise sexist behaviour	2.16	127	2.27	111	− 0.95	236	0.34	0.13
Prejudice	5.38	127	5.28	111	0.66	236	0.51	− 0.09
Equality	5.94	127	5.82	110	1.00	235	0.32	− 0.13
Tolerance of diversity in workgroups	4.11	126	4.10	110	0.12	234	0.91	− 0.02

Post-Intervention Results (Follow-Up 1)

Unpaired two-tailed t tests were used for all constructs to assess any differences between experimental and control groups at the first follow-up measure immediately following the V4V training. In total, 233 surveys were completed in the survey immediately after the V4V interventions (follow-up 1), of which 108 were from the experimental group and 125 from the control group, for an overall response rate of 86%, 82% in the experimental group and 98% in the control group. Table 3 presents results comparing the experimental and control groups at follow-up 1.

On four out of five constructs, Table 3 showed more desirable attitudes among recruits receiving the V4V programme compared to controls. The small effect sizes were not statistically significant ($p < 0.05$) except for tolerance of diversity in workgroups, the outcome on which the programme had the biggest effect at this stage.

The diversity construct asked participants about how they felt about interacting with people from different backgrounds in the workplace and how stimulated they felt in diverse workplaces. The mean score, out of a maximum of 5, was 3.87 for the control group compared to a mean of 4.10 for the experimental. The difference was not only significant, but "moderately" large as Cohen labels it—which is actually quite rare in much police research.

Table 3 Data and statistical analysis of constructs in survey post-intervention (follow-up 1)

Constructs	Control mean (score)	N	Experiment mean (score)	N	t	p	Cohen's d
Recognise racist behaviour	2.04	117	2.26	108	− 1.72	0.09	0.24
Recognise sexist behaviour	2.03	116	2.20	108	− 1.37	0.17	0.19
Prejudice	5.58	124	5.59	107	− 0.08	0.93	0.01
Equality	5.77	124	5.98	107	− 1.68	0.09	0.23
Tolerance of diversity in workgroups	3.87	124	4.09	108	− 3.11	0.002	0.41

Six Weeks Post-Intervention (Follow-Up 2)

The second follow-up survey took place six weeks after the V4V intervention. A two-tailed *t* test was used for all constructs to assess whether or not the experimental and control groups were the same or different at the second follow-up period. In total, recruits completed 194 surveys in the second follow-up survey, of which 105 were from the experimental group and 89 from the control group. The response rate was 79 and 70%, respectively. Table 4 presents results comparing the experimental and control groups at follow-up 2.

At the second follow-up survey, all five of the tests favoured the V4V group over the control group. While only the tolerance for diversity in workgroups had a statistically significant difference, that difference was "moderate" and almost "large" by Cohen's (1977) effect size standards. The other four were both small and marginally significant, with a maximum 11% chance of a false positive if we consider the groups to be different.

Comparing Results Across Baseline, Post-Intervention (Follow-Up 1) and Six Weeks Post-Intervention (Follow-Up 2)

Recognise Racist Behaviour in the Workplace In this construct, the V4V pro-gramme did not increase the capacity of recruits in the experimental group to recognise racism in the workplace, but it had a clear effect: *it buffered against declining capacities observed in the control group score results* (Fig. 2). At baseline, there was no difference between both groups. At follow-up, 1 there was an increase in the mean score of the experimental group to 2.26, which was higher than the control group mean of 2.04, a marginally non-significant difference (*p* = 0.086). At follow-up 2, both the experimental and control groups declined in their capacity to recognise racist behaviour over time, but by very little from baseline for the V4V group (mean score V4V = 2.17 and control = 1.93, respectively; *p* = 0.11). The control group declined more over time relative to the experimental group. This suggests that while V4V-treated recruits changed little in recruit training, recruits without V4V got discernibly worse.

Recognise Sexist Behaviour In this construct, the same pattern is found of V4V "buffering" recruits against the business-as-usual trend towards less ability to recognise sexist behaviour. While the score results for both the experimental and control groups

Table 4 Data and statistical analysis of constructs in survey 6 weeks post-intervention (follow-up 2)

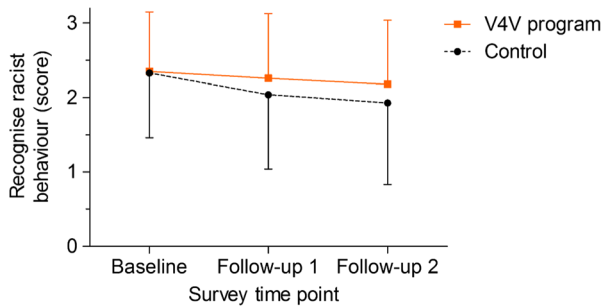| Constructs	Control mean *n* (score)	Experiment mean *n* (score)	*T*	*p*	Cohen's *d*
Recognise racist behaviour	1.93	73 2.18	101 − 1.60	0.11	0.26
Recognise sexist behaviour	1.89	73 2.13	101 − 1.59	0.11	0.25
Prejudice	5.08	89 5.35	103 − 1.62	0.11	0.26
Equality	5.49	89 5.74	103 − 2.02	0.05	0.29
Tolerance for diversity in workgroups	3.64	88 3.97	103 − 4.06	< 0.001	0.60

Fig. 2 Survey responses in the "recognise racist behaviour" construct before and after V4V participation. The data are presented as the mean score ± standard deviation for each group at each time point

over the duration of the study showed no significant differences, and both groups declined in their capacity to recognise sexist behaviour over time, the *control group declined more over time* relative to the experimental group (Fig. 3). This suggests that the V4V programme did not increase the capacity of experimental recruits to recognise sexism but rather safeguarded against the steeper declining views observed in the control group.

Prejudice Figure 4 shows that by the last survey on the prejudice construct, the V4V group had retained its initial level while the control declined, *with yet another buffering effect* of the programme. While there was little difference between the experimental and control groups over time in their capacity to recognise prejudices in the workplace, that difference supported the consistent pattern of less decay in the V4V group.

Equality Figure 5 shows *even clearer evidence of a buffering effect against an apparent values change in the course of recruit training*. Across the three survey waves, there was a decline in the control group's belief in equality from baseline to first follow-up and then in the second follow-up. By contrast, the experimental group increased from baseline to first follow-up in their beliefs concerning equality in the workplace, followed by a small decline by the second follow-up survey (5.82 at baseline, 6.00 at follow-up 1, and 5.74 at follow-up 2). By the second follow-up, the two groups were statistically significantly different in their views ($p = 0.05$; $d = 0.29$). Overall, the data suggest that in the experimental group, after an initial boost to their
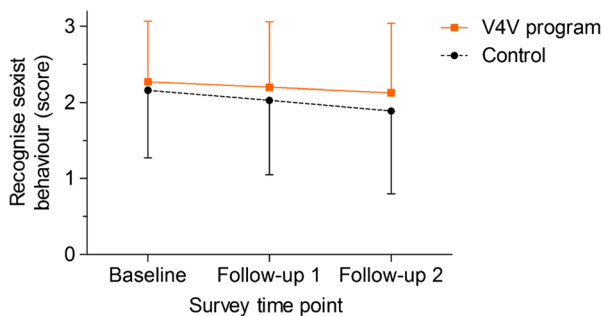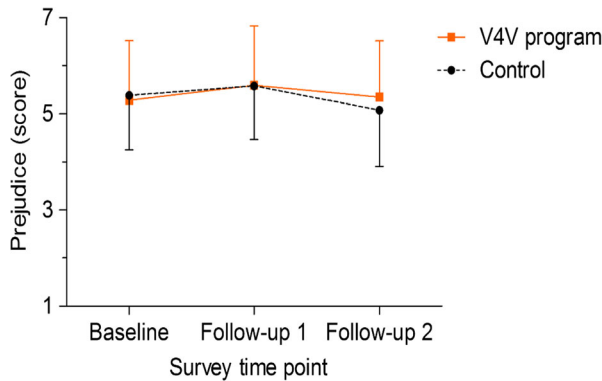

Fig. 3 Survey responses in the "recognise sexist behaviour" construct before and after V4V participation. The data are presented as the mean score ± standard deviation for each group at each time point
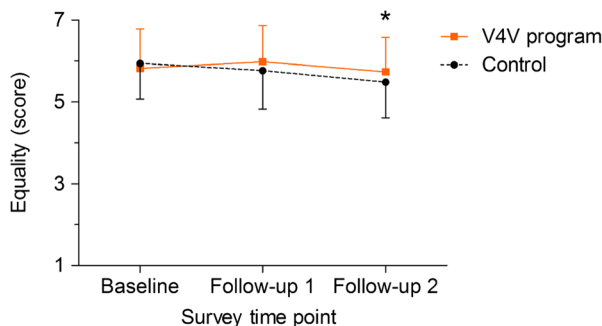
Fig. 4 Survey responses in the "prejudice" construct before and after V4V participation. The data are presented as the mean score ± standard deviation for each group at each time point

feelings about equality, there was decay in their views, suggesting that some of the V4V treatment effect could be short-lived as a booster of initially good attitudes, but long-lasting as a vaccine against the decline shown in the control group.

Tolerance of Diversity in Workplace This measure (Fig. 6) shows the *clearest evidence of all for a buffering effect of values education.* At both the first and second follow-up surveys, the V4V group showed significantly higher levels of tolerance for workplace diversity than the control group. Recruits who did not receive the V4V training demonstrated a decline of tolerance from baseline results (4.11 for control) to first follow-up results (3.87 for control) and second follow-up results (3.60). While both groups saw a decay in their tolerance towards workplace diversity, decay was so much greater in the control group relative to the experimental group that it caused the effect size to be the largest in this analysis ($d = 0.60$). The evidence is consistent with the interpretation that V4V protected recruits against declining tolerance for workplace diversity that appears to accompany the time spent in recruit training at the police academy.

Fig. 5 Survey responses in the "equality" construct before and after V4V participation. The data are presented as the mean score ± standard deviation for each group at each time point. Inter-group significant difference (*) at each time point was determined using a two-tailed t test ($p < 0.05$)
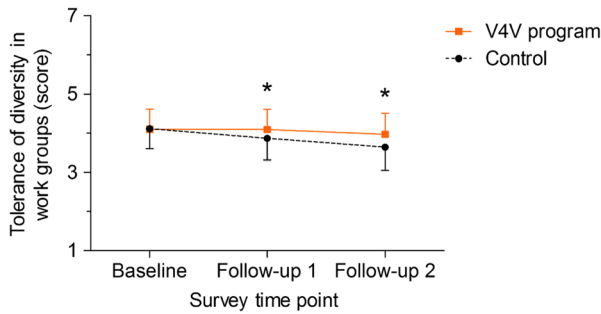
Fig. 6 Survey responses in the "tolerance of diversity in workgroups" construct before and after V4V participation. The data are presented as the mean score ± standard deviation for each group at each time point. Inter-group significant difference (*) at each time point was determined using a two-tailed *t* test (*p* < 0.05)

Summary

Despite the possibility of a "family wise error rate", across the different waves and constructs, eight constructs showed statistically significant differences between the experimental and control groups. Overall, the results show that the experimental group were more likely to voice pro-social values than the control group on their stated willingness to intervene in a racist incident (for follow-up 1 only) and stated willingness to intervene in a sexist incident (for follow-up 1 only). The experimental group also had a higher preference for equality (for follow-up 2 only), tolerance of diversity in workgroups (for both follow-ups 1 and 2) and empathy for discrimination (for follow-ups 1 and 2) than the control group. Whilst the results do not lead to sustained improvements in recruit attitudes and values across all constructs, the V4V programme clearly buffers against erosion of recruits' attitudes during the period that they attend the police academy as shown in the patterns of attitudes and values exhibited by the control group recruits.

Conclusion

During 2015, 132 police recruits at the QPS Academy received the V4V programme. Under randomised field trial conditions, the impact of the programme was evaluated to assess whether or not a values-based education programme at recruit level could increase recognition of racist and sexist behaviour, prejudice, empathy and tolerance of diversity. The answer to the research question was inconsistent.

On one hand the V4V programme did *not* increase recruit attitudes supporting the positive values taught in the course. On the other hand, the standard recruit training experience *reduced* support for those same values, absent the protective effects of V4V. Thus the main conclusion from this study is about tracking, not testing (Sherman 2013): the standard experience of Queensland police recruits in 2015 was to make them less tolerant of diversity than they had been before joining the police.

In the short run, the policy conclusion from this research may or may not be to roll out the V4V programme to every new recruit, or even post-recruit officers of all ages.

While there may be benefits for doing so, there may be even more benefits from replicating the experiment—not just in Queensland, but in every police agency in Australia, the UK and US. Similar values tracking research is underway at the police college in Denmark in conjunction with the Cambridge Police Executive Programme, but more randomised trials are clearly needed to increase the external validity of the results from the Queensland research.

Quite apart from the testing effects of the V4V programme, the more basic question comes from tracking recruit attitudes about diversity without the programme. Why did those attitudes show such clear declines? Was it caused by a peer culture that develops during the recruit course? Was it caused by implicit or explicit messages of the instructors? Was it caused by the way in which the recruits were treated by their friends and neighbours when the news spread about their appointment to the Police Service?

This study is unable to answer those questions. Yet the evidence the study reports shows how urgent those questions may be. Until there is a further and broader line of research completed, not just in Queensland but around the world, many police professionals will wonder what there is about police academies that may make them a major issue in the quest for police legitimacy.

References

Ashburn-Nardo, L., Morris, K., & Goodwin, S. (2008). 'The confronting prejudiced responses (CPR) model: applying CPR in organizations. *Academy of Management Learning and Education, 7*(3), 332–342.

Bachman, R., & Schutt, R. K. (2014). *The practice of research in criminology and criminal justice* (5th ed.). Los Angeles: Sage.

Bross, I. D. J. (1971). Critical levels, statistical language and scientific inference. In V. P. Godambe & Sprott (Eds.), *Foundations of statistical inference*. Toronto: Holt, Rinehart and Winston of Canada, Ltd..

Chan, J., Devery, C., & Doran, S. (2003). *Fair cop: learning the art of policing*. Toronto: University of Toronto.

Circelli, M. (1998). *Sexual harassment in the Queensland Police Service—a report on the outcomes of the NPRU 1998 'Experiences in the workplace' survey—pilot study*. Adelaide: National Police Research Unit.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciencies*. UK: Routledge.

Cox, T. H., & Blake, S. (1991). Managing cultural diversity: implications for organizational competitiveness. *Executive, 5*(3), 45–56.

Crime and Misconduct Commission (2010). Public perceptions of the Queensland Police Service. Findings from the 2010 Public Attitudes Survey. Download at www.ccc.qld.gov.au/public-perceptions

Cumming, E., Cumming, L., & Edell, L. (1965). Policemen as philosopher, guide and friend. *Soc Probl, 12*, 266–268.

Ellis, E., & Watson, P. (2012). *EU anti-discrimination law 2nd edition*. United Kingdom: Oxford University Press.

Ford, R. E. (2003). Saying one thing, meaning another: the role of parables in police training. *Police Quarterly, 6*, 84–110.

Harris, R. (2004). "Stepping into other peoples' shoes": teaching and assessing empathy in the secondary history curriculum. *International Journal of Historical Learning, Teaching and Research, 4(2) July 2004 ISSN 1472–9466 Editorial Board 2*, 98.

Kleg, M. (1993). *Hate prejudice and racism*. United States of America: New York Press.

McLeod, S. A. (2008). Prejudice and discrimination. Retrieved from www.simplypsychology.org/prejudice. html on 5 November 2015.

McNamara, J. H. (1967). Uncertainties in police work: the relevance of recruit's background and training. In D. Bordua (Ed.), *The police: six sociological essays*. New York: John Willey.

Nakui, T., Paulus, P. B., & Van Der Zee, K. I. (2011). The role of attitudes in reactions toward diversity in workgroups. *J Appl Soc Psychol, 41*, 2327–2351.

Nelson, J. Dunn, K. Paradies, Y. Pedersen, A. Sharpe, S. Hynes, M. and Guerin, B (2010). Review of bystander approaches in support of preventing race based discrimination.

O'Brien, M. (2011). Equality and fairness: linking social justice and social work practice. *J Soc Work, 11*(2), 143–158.

Pennay, D., & Paradies, Y. (2011). *The role of bystander knowledge, attitudes and behaviours in preventing race-based discrimination. A technical report on the conduct and findings of VicHealth's survey of bystander knowledge, attitudes and behaviors in preventing race-based discrimination*. Melbourne: Victorian Health Promotion Foundation.

Pennay, D. W., & Powell, A. (2012). *The role of bystander knowledge, attitudes and behaviours in preventing violence against women: a full technical report*. Melbourne: The Social Research Centre.

Platz, D. (2016). *The impact of a value education programme in a police recruit training academy: a randomised controlled trial*. Thesis submitted for the degree of Master's in Applied Criminology and Police Management, Institute of Criminology, University of Cambridge.

Polanin, J. R., Espelage, D. L., & Pigott, T. D. (2012). A meta-analysis of school-based bullying prevention programs' effects on bystander intervention behavior. *Sch Psychol Rev, 41*, 47–65.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: a personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67*(4), 741–763.

Sherman, L. W. (1980). Causes of police behavior: the current state of qualitative research. *J Res Crime Delinq, 17*, 69–100.

Sherman, L. W. (2013). The rise of evidence-based policing: targeting, testing and tracking. *Crime and Justice, 42*, 377–451.

VicHealth (2009) Building on our strengths: a framework to reduce race-based discrimination and support diversity in Victoria.

Victorian Equal Opportunity and Human Rights Commission (2015). Jenkins K independent review into sex discrimination and sexual harassment, including predatory behaviour, in Victoria Police retrieved from www.humanrights.vic.gov.au/victoriapolicereview.

Weisburd, D. (2010). Justifying the use of non-experimental methods and disqualifying the use of randomized controlled trials: challenging folklore in evaluation research in crime and justice. *J Exp Criminol, 6*, 209–222.

Deborah Platz MSt (Cantab) is Assistant Commissioner of the Australian Federal Police.

Elise Sargeant Ph.D. (Queensland) is Lecturer in the School of Criminology and Criminal Justice and member of the Griffith Criminology Institute at Griffith University, Brisbane, Australia.

Heather Strang Ph.D. (ANU) is Director of the Police Executive Programme at the University of Cambridge, Institute of Criminology, and was the academic supervisor of the Master's thesis research on which this article is based.