Demand management in the public sector: developments and issues

Paul Walley (paul.walley@open.ac.uk)
Open University, UK

Anna Jennison-Phillips
Gloucestershire Police, UK

Abstract

A study of the literature highlights that public sector organisations are developing more sophisticated practices to manage demand, and hence limit the total cost of service provision. These practices differ from those used by for-profit organisations. This paper studies demand management in the police service, drawing on the practices within five police force regions. The findings show improved understanding of demand and a focus on better processes to deliver services. The study develops a new framework for public sector capacity and demand management practices, contrasting revenue-centric and cost-centric approaches. This shows how the perspectives of demand management are fundamentally different.

Keywords: Demand, Capacity, Public, Police

Introduction

All public services sectors are under significant financial pressure, especially after the financial crisis from 2007. The UK Government spends approximately £780bn on public services (ONS, 2016) and welfare, but there has been a series of financial constraints applied to public expenditure, with many organisations or departments experiencing real-term cuts to budgets (e.g. HMIC, 2011). The situation is often made worse by increasing demand without corresponding increasing revenues. The challenge for managers within any of these services is to identify ways to *sustainably* reduce costs. Many organisations have looked used lean or quality improvement methodologies with mixed levels of success (e.g. Radnor & Osborne, 2013). An emerging alternative approach is to focus on the demand side of the problem. A key issue is whether or not demand can be limited to such an extent that resources can cope.

The operations management literature can assist in the analysis of this problem, through the existing body of demand and capacity literature (e.g. Sasser, 1976; Klassen & Rohleder, 2002). However, the idea of demand management appears to be less well-founded in a public sector context' partly because of the lack of scope to use pricing as a control mechanism in services that are free at the point of consumption (Walley, 2012).

1

This implies that public services must develop their own methods of managing demand to suit their own context.

The purpose of this study is to explore the current state of demand management practices in the public sector, to contrast these practices with those in the private sector and to highlight and evaluate any novel or previously unreported approaches to demand management that have been adopted by public sector organisations in response to resource constraints. The paper evaluates the literature based on both the public and private sector and also studies current attempts to manage demand in five police force regions of the UK. All police services in the UK are working at some level on demand management. The paper draws conclusions about the effectiveness of demand management practices in public services and derives a new model of demand management in operations that are not profit driven.

Demand and Capacity Management Literature

The origins of the demand management theory in services mainly emerges from the body of literature on medium-term capacity management. Johnston and Clark (2005) state that the task of capacity management is to achieve a balance between too much and too little resource utilisation. Service demand is commonly characterised as often instantaneous or unplanned, with seasonal and random variation (Lovelock 1992) making high utilisation a challenge. Many services also have degrees of variability (Frei, 2006) that not only make responsiveness and flexibility necessary characteristics of service operations, but also can lead to difficulties in planning capacity. There are two primary forms of capacity management defined in the literature (Sasser, 1976; Lovelock, 1992):

1. Chase capacity strategy where capacity is adjusted to meet short-term and seasonal demand fluctuations.
2. Level capacity strategies where capacity is mostly fixed and demand is managed so that peaks and troughs in demand are smoothed as much as possible, to better fit the capacity profile.

Fitzsimmons and Fitzsimmons (2004) identify five types of demand management:

1. Reservation systems and overbooking
2. Offering complementary services that deflect demand from the core service
3. Segmenting demand so that level capacity can be better utilised by using it to deliver counter-cyclical services
4. Price incentives and price manipulation to influence the size and timing of peaks and troughs in demand
5. Promotion for off-peak times, to deflect demand away from the peak.

Yield management (e.g. Kimes, 1989; Kimes and Chase, 1998) attempts optimise capacity utilisation in the short-term using both demand and capacity adjustment methods. Arguably, the literature has concentrated on more demand management as a primary method of maintaining the efficiency of service delivery (e.g. Rhyne, 1988; Radas and Shugan, 1998; Klassen and Rohleder, 2001), rather than looking at the ways in which the resources are controlled.

Public Sector Demand Management

At the present time there is limited literature that focuses purely on public services and their approach to capacity management. Walley (2012) identifies a contrast between public and private sector operations in the way they perceive the demand and capacity problem. Whereas in the private sector the planning approach starts with demand and goes through a series of steps to determine the capacity requirements, in the public sector the process is often in reverse. Public sector operations are often constrained by budgets that are set independently of demand (which may not even be counted accurately) and so the operations may have to determine what proportion of the total demand they are able to meet. This may also set off a series of activities associated with the prioritisation of work coming in, to postpone or reject new demand, because of existing capacity constraints. The most obvious examples of this would be in healthcare where triage or other prioritisation activities repeatedly grade work to always take the most urgent first.

One of the consequences of prioritisation is that variability of demand is often increased if it is compartmentalised into multiple types of demand and resource ring-fenced (carved out). In healthcare a considerable body of literature exists that explain the unnecessary queues and work backlogs created by such variability (e.g. Silvester et al., 2004). This consequence is observed as feeling counter-intuitive to operational managers, who, noticing more demand than they can effectively service, tend to prefer to put processes in place to determine and meet the most urgent demand instead of dealing with all "today's work today".

The newest literature, and the only literature found by the authors that directly confronts the challenge of demand management in the public sector, comes from the Local Government sector. The work takes a systems thinking approach building on the work of John Seddon. Seddon offers a systems methodology to improve service productivity. He introduced the idea of failure demand, which can be defined as "demand caused by a failure to do something or do something right for the customer" (Seddon, 2003). When discussing service improvement he makes a bold claim:

> *"In service organisations…failure demand often represents the greatest lever for performance improvement. In financial services it can account for anything from 20 to 60 per cent of all customer demand…in local authorities and police forces as much as 80-90% are avoidable and unnecessary"*
>
> Seddon, 2009, p33

This statement implies that most public services should easily be able to meet all true demand if they are able to eliminate the failure demand from within their systems – and the UK government agreed to implement such an approach in 2008. The first attempt to implement approaches to failure demand reduction was to impose a performance measure, NI14, for the measurement of *avoidable demand*. The Government advice defined avoidable contact in several ways (IDEA, 2008):

1. Unnecessary clarification by the customer
2. Incorrect contact, such as a customer being put through to the wrong department
3. Repeat contact, such as a customer having to inform separate council departments of a change in address
4. Customers having to progress-chase work
5. Repeat contact due to work not finished

At the time, some concerns were raised about the difficulties of implementing the measure effectively. One study (Olsen, 2008) listed a number of potential barriers to its implementation, such as the lack of senior stakeholder support, lack of capacity or skills internally to deal with the data and lack of guidance. The measure was not implemented for long. In March 2010 it was discreetly removed from the Government's list of measures to be reported (see Martin, 2010).

More recently there has been an attempt by the Local Government Association, amongst others, to reinvigorate the application of failure demand concepts in the public sector. They take a systems approach to demand reduction, extending the failure demand model to include other types of unwanted demand. Table 1 provides a summary of these.

Table 1 Types of Unwanted Demand (Randle & Kippin, 2014)

Type of demand	Explanation
Failure	Demand caused by errors or poor processes
Avoidable	Demand arising from behaviours that can be changed
Excess	Demand created by providing more than is needed
Preventable	Demand arising from causes that could be removed earlier
Co-dependent	Demand that is unintentionally reinforced by dependence

The first of these types of unwanted demand is what might be described as conventional failure demand. Avoidable demand is often generated by routines where service is provided irrespective of need. Some services have specifications that exceed minimum requirements and service users could be satisfied with less service. Co-dependent demand is often repeat demand where a service has been provided and now the user relies on this service on an ongoing basis. Randle and Kippin (2014) also make suggestions for how demand can be prevented. Table 2 shows some of the suggested generic approaches:

Table 2 Actions suggested to reduce demand

Demand	Methods of prevention
Failure	Service redesign Quality improvement
Avoidable	Changing employee behaviours Changing relationships with customers Shifting towards prevention
Excess	Charging Punitive measures for non-compliance Changing eligibility criteria
Preventable	Understanding and tackling root causes Understanding behavioural influences
Co-dependent	Building community resilience Alternative strategies for community or individual capacity

These authors present a series of examples of how demand for local public services have been reduced. For example, some councils are introducing further charging for bulky or green waste removal. Other councils are changing eligibility criteria for adult social care. For instance: police services make a charge for responses to Freedom of Information

requests; the Fire and Rescue service has been very successful at dealing with preventable causes of fires.

Research Methodology
This study focuses on an area that is not widely covered in the existing literature – the police service. An initial scoping study was conducted in the Thames Valley region, where 10 police employees, ranging from community support officers, sergeants, inspectors, to police staff were given structured interviews to establish what demands were placed on their time for core policing work and administration. This study then informed the approach taken in "evidence cafes", based on the method used by Café Scientifique. The cafes were attended by 25-40 police officers and police staff from all levels of each organisation. A short discussion of demand and capacity methodologies, including an explanation of failure demand, was followed by a discussion in groups of the applicability of the ideas to the police. Questions asked included:
1. What are the main sources of unnecessary demand?
2. Are these sources internal or external?
3. What services do the public want that should not be provided by the police?
4. What services are not currently provided that are needed?

Part of the session also used "Q-board" method to identify perceptions of value of police activities (Stephenson, 1953; Watts and Stenner, 2005). This was conducted in three police regions: Humberside, Hertfordshire and Bedfordshire. All sessions were recorded and data taken during review of these transcripts.

Findings
Table 3 shows examples of the opinions expressed for the different types of avoidable demand suggested by the research.

Table 3 Suggestions for police avoidable demand

Demand	Examples
Failure	Attendance where victim does not want to report crime
	Repeat unnecessary contact
	Defensive approach to risk assessment creates extra work
Avoidable	Mandatory despatch of officers to high risk "mispers"
	Shed burglaries
Excess	Attendance at non-injury road traffic accidents
	Attendance for drunk people who are injured
	Social media bullying, nuisance parking, noisy neighbours
Preventable	Shoplifting, especially from high street chains
	Theft from unlocked vehicles
Co-dependent	Calls from care home
	Attendance for other agencies

These findings were reinforced by the Q-board which highlighted some of the perspectives on the value work that police do.

5

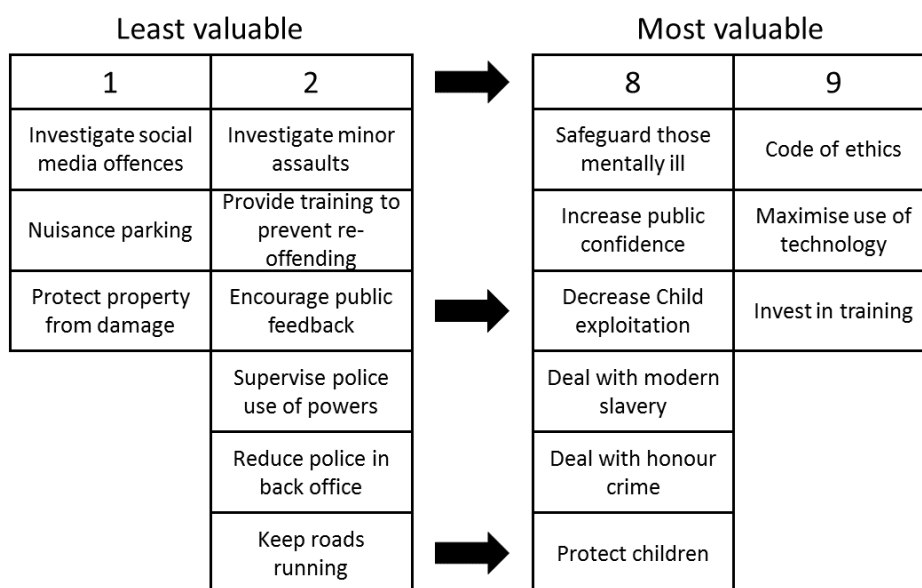| Least valuable			Most valuable	
1	2		8	9
Investigate social media offences	Investigate minor assaults		Safeguard those mentally ill	Code of ethics
Nuisance parking	Provide training to prevent re-offending		Increase public confidence	Maximise use of technology
Protect property from damage	Encourage public feedback		Decrease Child exploitation	Invest in training
	Supervise police use of powers		Deal with modern slavery	
	Reduce police in back office		Deal with honour crime	
	Keep roads running		Protect children	

Figure 1 Sample answer to Q-board exercise (extremes only)

The results show that some of the high volume sources of demand, such as calls due to cars parked blocking driveways or noisy neighbours, are not regarded as value work by police officers. It would be interesting to contrast these results with a tally of the subjects which most frequently appear in emails from members of the public to Police and Crime Commissioners.

Discussion & Contribution

One of the most striking aspects of the responses given is how provider-based, rather than user-based, the answers and discussions became. Much of the higher volume work is seen as low value by the police themselves, but the public expect to receive a service from "someone" to deal with these matters, in exchange for the taxes they have paid into the public sector. Superficially, at least, there was always a strong opinion that the police were better judges of where to devote their time than members of the public and this creates some tension when deciding what services to provide. There were also frequent discussions of how the public (and other agencies) misused the capacity the police did have.

   It becomes apparent that the perspective in this public service is cost-centric, rather than revenue centric, as few, if any, services offered generate revenue. Even services which generate revenue are considered with a cost-centric mindset. For example, Financial Investigators often generate more revenue from Proceeds of Crime Act proceedings than their own salaries, yet are considered a "nice to have" resource when times are lean. Likewise, fines from traffic offences can be more than sufficient to sustain a team of staff processing the offences, but the staffing and other resourcing of such units is not immune to cuts which effectively limit a revenue-generating demand from being serviced. In part this may be due to the revenue formulas applied whereby only a small proportion of the revenue can be returned to the employing police force, and in part it may be due to the public unpopularity of pursuing such punitive strategies; equally it could be a reflection of the wider police culture in which demand generally is deemed to be unwanted as it causes limited resources to be spread ever more thinly.

We would propose that a cost-centric demand management model is developed as a contrast to the revenue centric models currently used in the service management literature. The discussions during the evidence cafes revealed five broad categories of demand management that were currently "possible" for the police services:

1. Refuse to meet demand

There was a lot of discussion about the types of service that the police should refuse to provide, even where they currently provide some sort of service. Most of these examples emanate from the suggestion that other agencies, or the callers themselves, have responsibility for some types of demand. One example would include the problem of "noisy neighbours" which is a civil matter that should be dealt with by local government through the use of anti-social behaviour orders (ASBOs) where necessary. Some officers also suggested that some services could be limited so that when the police were short of resource a service should be routinely refused, for instance, theft of fuel from service stations.

2. Limit demand

Many mechanisms were suggested for limiting of demand. Existing prioritisation systems grade all new calls for risk, with all high risk incidents attended as soon as possible. Many calls are time sensitive and so a delay to a call may cause it to be downgraded, ultimately to one that does not need a response. For incidents of shoplifting there was a clears sense amongst some officers that not all demand should be met. High street stores, for example, have merchandising practices that make shoplifting easier and are financially robust enough to withstand the losses this creates. The police felt that this did not usually need a response. However, a small corner shop that experiences problems that are more financially and psychologically impactful was seen as worthy of a response.

Demand prevention, especially activities that reduce crime, featured in all discussions. Although there are some actions, such as the use of CCTV in high streets, that help reduce some crimes, the issue discussed was that much crime prevention is achieved through behaviour changes of potential victims, e.g. locking of cars at night, putting locks on sheds, etc.

Another observation was that some offenders (and callers) repeatedly created demand. An example would be a care home that frequently reported young persons missing at night. Much repeat demand has an underlying social root cause, such as drug misuse, and there was discussion of the underlying social root causes of demand. Some suggestions were that repeat demand could incur a charge, especially where the repeated demand was from another organisation which undoubtedly had more capability to prevent such demand recurring.

3. Move demand

Police officers consistently identified that much avoidable demand should be dealt with by other agencies. A common problem is demand created by those with drug or alcohol problems where anti-social behaviour or minor crime may be committed. Arguably, agencies with a remit to reduce substance addiction are likely to have a greater impact on the criminality of their clients than the police. Police are often tasked with minding those with mental illness as they can provide a so-called "place of safety". Similarly, there are many civil offences that are the responsibility of individuals or councils to deal with (such

as travellers parking caravans on private land, fly tipping etc.). Much of this demand comes into the police call centres instead – often with no mechanisms to pass this demand on efficiently. This is one area of activity in local government that has gained some traction as "partnership working".

4.   Change the service offering

One of the ideas that surfaced quickly in the Gloucestershire is the idea of developing a better array of services that could be offered, more tailored to suit the needs of the callers, making better use of developments in technology. Opportunities to use technology such as skype/facetime, email or other IT-related forms of communication are partly explored but are not part of standard procedure. These types of technology could prevent the need to attend some incidents, whilst providing the personal contact between the police and the public which is needed to reassure and advise.

There are also opportunities to make services either more self-service, or to reduce the need for intervention by public agencies. In the literature this is sometimes referred to as community resilience. One example uncovered during the study is when callers report "animals in the road". The callers could be encouraged to deal with the problem themselves on occasions when it is safe to do so.

5.   Reduce the resources needed

Within the public sector literature the concept of failure demand was a dominant feature of the discussion. The idea that most demand is unnecessary is appealing at one level. However, this idea seems not to have taken hold. Detailed analysis of some calls showed that there were examples of multiple contact for one incident where one contact may only have been needed if it was dealt with at the first attempt. As an example, in a shoplifting incident the police failed to respond within an hour and this resulted in three extra calls including two calls to say that the suspected perpetrators had returned to the same store.

The above discussion leads to the development of a cost centric model of demand management practices for services where revenue is not directly linked to the demand entering the system. This model is proposed as an extension of existing medium-term capacity management practice, shown in figure 2.

We accept that this model is phenomenological in that it reflects practice, but does not necessarily present the ideal model for not-for-profit demand management. We would suggest that the theory should include a more service-centric or value-centric component where the refusal for service or the limits to service are more explicitly based around objective measures of value, rather than situational constraints on capacity. At present this will remain a theoretical ideal, in a world where total resource allocation is based more on budgets than operational need.
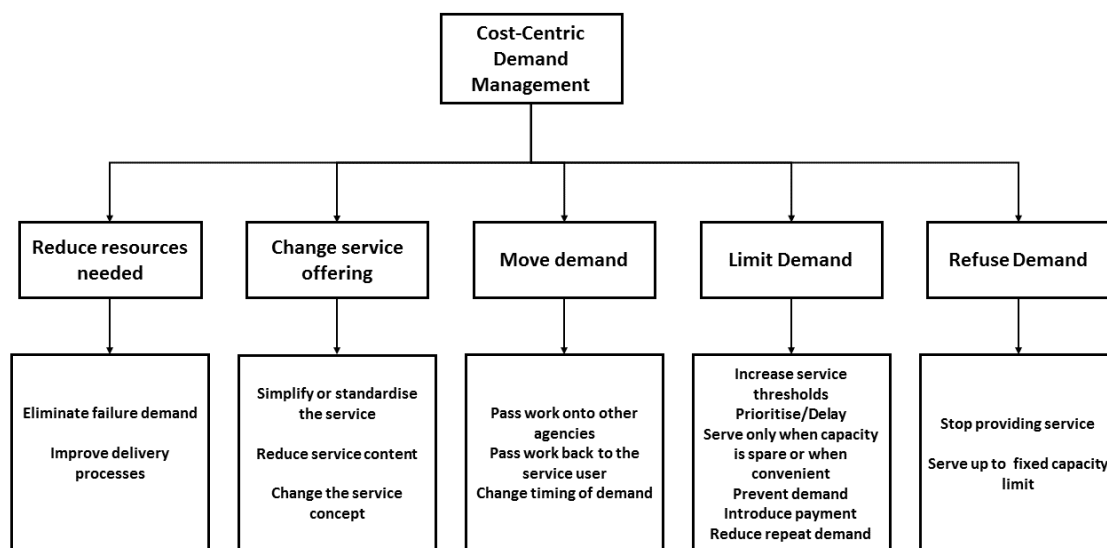
Figure 2 Cost-centric demand management practices

Conclusions

Work in the public sector in the UK shows new developments in the way demand is managed in services, especially where there are limited opportunities to charge directly for the services provided. The approach is based upon some sound theory from lean practices and systems thinking. This research shows that the ideas are, in principle, applicable to services beyond those in local government. However, our analysis shows there are a number of issues:

1. There can be tension between the service providers who dispute the value of providing a service and the service users who benefit from the provision of the service. In this case much of the high frequency demand is classed as low value by service providers.
2. Some of the mechanisms to deal with perceived excess demand are simply to either refuse to provide a service on the basis of lack of resource or to raise the risk thresholds of when service will be provided. This challenges the systems idea that a service should be designed to meet demand, however the longer term impact and sustainability of pursuing such a strategy has not been considered here. We would also question the extent of the freedom to refuse to provide a service in situations where resources may be enough to meet demand if it is efficiently used.
3. The lack of evidenced activity, so far, to reduce failure demand raises the question of why this has not been tackled sooner. It is likely there are many factors, similar to those listed in the literature for the failure of NI14, that prevent large-scale implementation currently.

Finally, this paper has presented a model of cost-centric demand management that we suggest can complement existing medium-term demand and capacity theory.

References

Frei, F.X., (2006), "Breaking the Trade-Off Between Efficiency and Service", *Harvard Business Review*, November, 93-101.

HMIC (2011) *Adapting to austerity*, HMIC, London.

IDEA (2008), *Reducing Avoidable Contact, a guide to N14*, Improvement and Development Agency, London.

Fitzsimmons, J.A. And Fitzsimmons, M.J. (2006) , *Service Management: Operations, Strategy, Information Technology*, New York, McGraw-Hill.

Jackson, M, Johnston, N and Seddon, J (2007), "Evaluating Systems Thinking in Housing" Journal of the Operational Research Society no 59, 186–197

Johnston, R. and Clark, G. (2005), *Service Operations Management*, Pearson Education Ltd Essex.

Kimes S.E. and Chase R.B. (1998), "The Strategic Levels of Yield Management", *Journal of Service Research*, 1(2), 156-166.

Kimes, S.E., Chase, R.B. and Wirtz, J. (2003), "Has Revenue management become acceptable? Findings from an International study on the perceived fairness of Rate Fences", *Journal of Service Research*, 6(2), 125-135.

Klassen, K.J And Rohleder, T.R. (2002), Demand and capacity management decisions in services; how they impact on one another, *International Journal of Operations and Production Management,* 22 (5), p527-548.

Klassen, K.J. and Rohleder, T.R. (2001), "Combining operations and marketing to manage capacity and demand in services", *Service Industries Journal*, 21(2), 1-30.

LGA (2013), *Managing Customer Demand: Understanding and changing behaviours to help meet the financial challenge,* LGA, London.

Martin, P, (2010), "Avoidable Contact: If only…", Local Government Chronicle, http://www.lgcplus.com/news/finance/efficiency/avoidable-contact-if-only/5013769.article (Accessed 22/06/2015)

Masters, K.I.A., (2009), *From New Public Management to Lean thinking: understanding and managing 'potentially avoidable failure induced demand*, Unpublished PhD Thesis, University of Surrey.

Olsen, H. (2008), "NI 14: Help or Hindrance?", http://www.ukauthority.com/About-Us/Research-Projects/NI14.aspx, last accessed 19/6/2015

ONS (2016), http://visual.ons.gov.uk/uk-perspectives-2016-spending-on-public-services-in-the-uk/ last accessed 10/01/17

Radas, S. and Shugan, S. (1998), "Managing Service Demand: Shifting and Bundling", *Journal of Service Research*, 1:1, 47-64.

Radnor, Z. and Osborne, S.P. (2013), "Lean: A failed theory for public services?", Journal of Public Management Review, Vol. 15, No. 2, 265-287.

Randle, A. and Kippin, H. (2014), *Managing Demand: Building future public services*. RSA, London.

Rhyne, D.M. (1988), "The Impact of Demand Management on Service System Performance", *Service Industries Journal*, 8(4), 446-458.

Sasser, W.E. Jr (1976), Match Supply and Demand in Service Industries, *Harvard Business Review*, 54 (6), p133-140.

Seddon, J (2003), *Freedom from Command and Control*, Vanguard Press: Buckingham.

Seddon, J (2009), "Failure Demand – from the horse's mouth" Customer Strategy Issue 1, Vol 2 Winter 2009 pp 33-34

Silvester, K, Steyn, R, Bevan, H, And Walley P. (2004), "Reducing waiting times in the NHS: is lack of capacity the problem?", *Clinician in Management,* 12, p105-111.

Stephenson, W. (1953), *The study of behaviour: Q techniques and its methodology*, University of Chicago Press, Chicago, Il.

Walley, P. (2012), "Does the public sector need a more demand-driven approach to capacity management?", *Production Planning and Control*, Online March 2012, ref: DOI: 10.180/09537287.2012. 2012666886, pp1-14

Watts, S. and Stenner, P. (2005), "Doing Q methodology: theory, method and interpretation", *Qualitative Research in Psychology*, No. 2, pp67-91.