



## ประเมินผลงานด้วยโพลนั้นเชื่อถือไม่ได้

วารสารธรรม สามโกเศศ

วันที่ 15 ตุลาคม พ.ศ. 2552

หน้า 6

การที่โพลระบุว่าประชาชนประเมินผลงาน

นายกรัฐมนตรีหรือรัฐมนตรีคนใดว่าได้คะแนน 4 จาก 10 นั้น ถ้าพิจารณาให้ดีแล้วมันแทบไม่มีความหมายอะไรเลย เช่นเดียวกับผลจากโพลในเรื่องอื่นๆ ในบางครั้ง โพลนั้นจะมีประโยชน์ก็ต่อเมื่อผู้ทำโพลถามคำถามและจัดให้มีการประเมินอย่างมีความหมายและประชาชนผู้รับสารเข้าใจข้อจำกัดของผลจากโพลเหล่านั้นเท่านั้น

ในเบื้องต้นถ้ากลุ่มที่ถูกสำรวจไม่ได้เป็นตัวแทนของประชาชนทั้งหมดดังที่โพลคาดหวัง ผลก็ผิดพลาดแต่แรกอุปมาถ้ำน้ำแกงในช้อนที่ตักมามีได้เป็นตัวแทนของหม้อแกงทั้งหม้อ รสชาติของแกงในช้อนก็ไม่ได้มีความหมายอะไรเลย

การคัดกลุ่มตัวอย่างของการสำรวจก่อนทำโพลจึงเป็นเรื่องสำคัญยิ่ง ถ้าต้องการรู้ว่าคนไทยทั่วประเทศรู้สึกอย่างไร กลุ่มตัวอย่างก็ต้องเป็นตัวแทนของประชาชนทั้งประเทศ กล่าวคือองค์ประกอบของกลุ่มตัวอย่างต้องเหมือนองค์ประกอบของคนไทยทั่วประเทศ

การเอาแบบสอบถามออกไปถามกลุ่มตัวอย่างที่มีลักษณะดังต้องการไว้ก็ต้องกระทำด้วยความระมัดระวังอย่างยิ่งประเด็นสำคัญก็คือ "การนั่งเทียน" ของผู้ถาม ถ้าไม่ควบคุมให้ดี ผู้ถามก็จะตอบเสียเอง เพราะชี้แจงออกไปเดินท่อมๆ ตากแดดตามหาสมาชิกกลุ่มตัวอย่างที่มีลักษณะตามที่ปรารถนาเพื่อให้ออกคำถาม

ถึงแม้จะพยายามใช้วิธีคัดกรองทางวิชาการ เพื่อให้ได้ข้อมูลที่ไม่ได้มาจากการ "นั่งเทียน" เพียงใด ก็ไม่อาจหลีกเลี่ยง "พลังเทียนไข" ไปได้ทั้งหมด หากแบบสอบถามมิได้ออกแบบเพื่อให้สามารถตรวจสอบการ "นั่งเทียน" ได้อย่างมีประสิทธิภาพพอควรแล้ว ผลก็จะยิ่งไม่มีความน่าเชื่อถือ

ถ้าให้ผู้ถามคำถามความเห็นสามารถเอาแบบสอบถามกลับบ้านข้ามคืนได้แล้ว พลัง "เทียนไข" ก็จะสูงยิ่งขึ้น ปัจจุบันมีการถามความเห็นทางโทรศัพท์จากกลุ่มตัวอย่างที่ได้คัดสรรไว้แล้ว ดังนั้น "พลังเทียนไข" ก็น้อยลงไปแต่ก็ไม่หมด เพราะสมาชิกที่ไม่รับผิดชอบอาจให้คนอื่นตอบคำถามแทนหรือตอบอย่างพอไปทีเพราะได้รับผลตอบแทนแล้วจากการเป็นกลุ่มตัวอย่าง

สมมุติว่าได้กวนหม้อแกงเป็นอย่างดีและตักออกมาหนึ่งช้อน โดยแน่ใจว่าเป็นตัวอย่างน้ำแกงในหม้อแกง ปัญหาที่ยังไม่จบ คำถามและลักษณะการให้คำตอบของโพลเป็นสิ่งสำคัญอย่างยิ่ง

การให้ประชาชนประเมินผลงานโดยให้สเกล 1 ถึง 10 และเมื่อได้ผลมาแล้วก็เอามาเฉลี่ยและดูว่าเกิน 5 ซึ่งหมายถึงว่า "สอบได้" หรือไม่ นั้นเป็นเรื่องที่ผิดพลาดเพราะประการแรกผู้ตอบแต่ละคนมีมาตรฐานของการประเมินไม่เหมือนกัน บางคนเข้มงวดมากถึงจะทำงานดีก็ให้เพียง 3 บางคนถึงทำงานแย่ก็ให้ 4 ฯลฯ เมื่อมีเกณฑ์ "ผ่าน" ที่แตกต่างกัน ดังนั้น การเอามาบวกกันและหารเฉลี่ยและเอา 5 เป็นเกณฑ์ผ่าน จึงไม่มี ความหมาย

ประการที่สอง เมื่อสเกลคะแนนกว้างเช่นนี้ การให้คะแนนจึงสะเปะสะปะเพราะไม่อาจตอบได้ว่า 9 คะแนน กับ 7 คะแนนหรือ 6 กับ 8 ซึ่งต่างกันถึง 2 คะแนน นั้นมีเหตุผลมาจากอะไร

ประการที่สาม แต่ละคนให้ความสำคัญแก่คะแนนแต่ละระดับไม่เท่ากัน สำหรับบางคนพอใจเต็มที่ให้ 10 คะแนน บางคนถึงเต็มใจเต็มที่อย่างดีก็ให้แค่ 7 บางคนไม่พอใจก็ให้ 4 คะแนน บางคนไม่พอใจก็ให้ 0 คะแนน ฯลฯ

ในความเห็นผู้เขียน ถ้าจะช่วยให้การประเมินเช่นนี้มี ความหมาย น่าจะมีสเกลที่ให้เพียง 1 ถึง 5 และมีคำอธิบายกำกับว่าแต่ละระดับคะแนนที่ให้ความหมายอย่างไร เช่น ระบุว่าให้ 3 เท่ากับพอใจปานกลาง ส่วน 4 คะแนนนั้นหมายถึงสูงกว่าพอใจปานกลางแต่ยัง

ไม่พอใจที่สุด ส่วน 5 นั้นพอใจสูงสุด ในด้านคะแนนน้อยก็ให้คำอธิบายว่า 2 คือ ต่ำกว่าพอใจปานกลางแต่ยังไม่ถึงไม่พอใจสูงสุด ส่วน 1 นั้นคือ ไม่พอใจที่สุด

การให้ประเมินโดยสเกลเช่นนี้ จะเป็นการบังคับให้ประเมินอย่างชนิดที่แก้ไขปัญหา 3 ข้อข้างต้นได้ และสามารถพอจะเอามาเฉลี่ยกันและมีความหมายได้ สมมุติ นายกรัฐมนตรีได้คะแนนความพอใจออกมา 3.5 ก็หมายถึงว่าโดยเฉลี่ยกลุ่มตัวอย่างซึ่งเป็นตัวแทนของประชาชนไทย มีความพอใจการทำงานสูงกว่าปานกลาง แต่ไม่ถึงระดับพอใจมากที่เป็นระดับ 4

ในการตีความโพลต้องทำให้ประชาชนเข้าใจว่า "3.5" นี้คือคะแนนเฉลี่ย ซึ่งหมายถึงว่ามีคนจำนวนหนึ่งที่ประเมินให้คะแนนสูงกว่า 3.5 และจำนวนหนึ่งที่ประเมินให้คะแนนต่ำกว่า 3.5 (เปรียบเทียบในห้องหนึ่งคนมีความสูงเฉลี่ย 1.65 เมตร มิได้หมายความว่า จะมีคนสูง 1.65 เมตรอยู่ในห้องนั้นหลายคน อาจไม่มีคนสูง 1.65 เมตรสักคนอยู่เลยก็ได้ จำนวนหนึ่งอาจเป็นปิกมี (Pigmy) จากนิวกินี ซึ่งสูงเพียง 1 เมตรหรือ 1 เมตรกว่า และอีกจำนวนหนึ่งอาจเป็นคนเผ่ามาไซ (Maasai จากแอฟริกา) ซึ่งสูง 2 เมตรกับ 2 เมตรกว่า ตัวเลข 1.65 เมตรมาจากความสูงเฉลี่ยของคนทั้งหมดในห้อง)

ถ้าจะให้โพลมีความหมายมากกว่านี้ ต้องบอกด้วยว่าความพอใจการทำงานของนายกรัฐมนตรีมีลักษณะคะแนนอย่างไร มีร้อยละเท่าใดที่ประเมินให้คะแนน 5 4 หรือ 2 หรือ 1 มิใช่การให้คะแนนประเมิน 3 หรือ 4 หรือ 5 จากคะแนนเต็ม 10 ซึ่งแทบจะไม่มีความหมายเอาเลย เมื่อคำนึงถึงว่าคะแนนเหล่านี้มาจากคนที่รู้เรื่องต่างๆ ไม่เท่ากัน แต่นับเป็นหนึ่งเท่ากันในกลุ่มตัวอย่างแล้วยิ่งรู้ว่าโพลมีข้อจำกัด

การประเมินความพอใจผ่านการสำรวจในต่างประเทศเท่าที่เคยเห็นก็ไม่มีที่ให้คะแนนเต็ม 10 (เห็นมีแต่มวยสากลที่ให้แต่ละยก ซึ่งโดยประเพณีก็มีการให้กันแค่ 10 หรือ 9 คือถ้าใครชนะยกนั้นก็ได้ 10 หากอีกคนแพ้ก็จะได้ 9 และถ้าแพ้ยับเหินก็จะได้ 8 คะแนน กรณีนี้ผู้ให้คะแนนมีเกณฑ์การให้แต่ละคะแนนที่ชัดเจน การใช้คะแนนเต็ม 10 หรือ 3 จึงไม่แตกต่างกัน)

สิ่งที่รู้สึกรับชั่งและไม่เคยเข้าใจก็คือมีผู้ส่ง SMS โหวตลงคะแนนโดยยอมเสียเงิน 3 บาท หรือมากกว่านั้นกัน ในการตัดสินการแข่งขันร้องเพลง AF (ซึ่งก็คือโพลอย่างยิ่ง) การใช้คะแนนโหวตแบบนี้เพื่อวัดความนิยมโดยคนดู

ที่บ้านนั้นผิดหลักสถิติของการตัดสินความนิยม ผู้แข่งขันคนใดที่มีผู้สนับสนุนที่มีเงินมากก็ได้คะแนนมากอย่างบิดเบี้ยวไปจากความเป็นจริงของความนิยม ที่ซับซ้อนก็คือไม่มีทางรู้ว่าโหวตที่ตนเองยอมเสียเงินค่า SMS นั้น ได้รับการนับเข้าไปหรือเปล่า เพราะไม่เคยมีการประกาศคะแนนโหวตให้รู้ ทุกอย่างลึกลับ ไม่โปร่งใสและไม่เป็นธรรมแก่ผู้บริโภค (คนดูเสียเวลาและค่าไฟฟ้าดูและมีคนเสียเงินส่ง SMS แต่ไม่รู้ว่าผลการโหวตจริงเป็นเช่นใด)

ความเห็นจากโพลมักถือกันว่าเป็น **public opinion** (ความเห็นของสาธารณชนหรือบางคนไปไกลถึงกับบอกว่า เป็นมติมหาชน) แต่เนื่องด้วยสารพัดข้อจำกัดของโพลจริงๆ แล้วมันจึงไม่มี **public opinion** ที่แท้จริง

จะมีก็แต่เพียง **opinion in public** หรือความเห็นที่ปรากฏให้เห็นในสาธารณะเท่านั้น