

A Nationwide and Temporal Study of Learning Internet Queries in the United States

Dr. Jun Tsuji

Siena Heights University
1247 E. Siena Heights Drive
Adrian, Michigan 49221 USA

Abstract

The internet is an important source of information for learning. Consequently, web searches for the term “learning” may serve as an indirect indicator of learning. This hypothesis was assessed using 5-years of Google search engine query data for the U.S. population. “Learning” web searches exhibited seasonal patterns of activity that mirrored the academic calendars of U.S. schools and colleges, were markedly greater on days when students were in attendance at school, and were positively correlated with web searches for math-related content areas. Google search engine queries for “learning” were also positively correlated with internet searches for “motivation,” as predicted by the Unified Learning Model and other learning frameworks. These observations suggest that “learning” query data can be a reasonable proxy for monitoring seasonal learning.

Keywords: learning, motivation, internet search, web query, Google trends

Introduction

In the United States of America, the internet is the primary source of information for most people (Jansen and Spink 2006). Over the past decades, an ever increasing quantity of educational content has been uploaded on the World Wide Web. For instance, schools are using Blackboard electronic learning management system to post classroom assignments and assessments online (Falvo and Johnson 2007), millions of tutorial videos can be viewed on YouTube (Cheng et al. 2008), and hundreds of thousands of library books and reference materials can be accessed through the Google Books Library Project (Jones 2011). These web sites and others provide abundant sources of subject material for learning.

Traditional measures of learning utilize tests (Amrein and Berliner 2002), surveys (Kuh 2009), and behaviors (Angelo and Cross 1993) that are limited to a specific population and/or to a particular moment in time. While informative, such approaches limit our ability to describe trends in learning on a nationwide and temporal scale. Knowing such breadth of scale would better inform our understanding of learning and in turn may influence teaching practices and educational policies. In an attempt to expand measures beyond those of traditional methods, I explored an approach based on internet search engine queries for the term *learning*.

Individuals seeking knowledge can utilize internet search engines, like Google, to locate information relevant to their query (Tene2007). The history of millions of such web searches is publicly available through the online tool Google Trends (Shimshoni et al. 2009). Google Trends utilizes an algorithm to calculate the number of search queries for a given term within a specified geographical region and timeframe and then divides the sum by the total number of queries for the same location and time period (Choi and Varian 2012). The ratio is then scaled (0-100) relative to the week with the greatest number of web searches (Desai et al. 2012a). Research on search engine query data has revealed that web search activity can be indicative of specific populations and can occur in defined patterns (Ettredge et al. 2005; Polgreen et al. 2008; Shimshoni et al. 2009; Ginsberg et al. 2009; Hulth et al. 2009; Desai et al. 2012a; Desai et al. 2012b; Bakker et al 2016). For instance, job-related web searches reflect the number of unemployed workers (Ettredge et al. 2005), and web search queries containing the words *flu* or *chicken pox* correlate with the number of influenza and chicken pox cases, respectively (Ginsberg et al. 2009; Bakker et al. 2016). As job and medical web queries provide information on employment and disease trends, I hypothesized that web searches for the term *learning* may serve as an indirect indicator of learning. This paper describes a nationwide and temporal study of *learning* internet queries in the United States.

Methods

Weekly Google search data for the entire United States from November 6 2011 to November 6 2016 was collected for the terms *learning*, *motivation*, *math*, *algebra*, *quadratic*, *exercise*, *workout*, *fitness*, and *diet* from Google Trends (2016). Daily *learning* web search activity for the U.S. population from September 6 2016 to November 6 2016 was also examined. The daily *learning* Google search queries in the state of Rhode Island were compared to the daily public and private school kindergarten-grade 12 student attendance numbers reported by the Rhode Island Department of Education (2016) from September 6 2016 to November 6 2016. Daily student attendance data for other states were not publicly available. Linear correlations between search term queries were assessed using coefficients of determination (Bakker et al. 2016).

Results and Discussion

Google web searches for the term *learning* exhibit temporal patterns, which mirror the academic calendars of United States kindergarten - grade 12 schools and colleges. Over the course of the 5-year study period, weekly *learning* web searches increased at the end of August/early September with the start of the academic year. Web search activity then exhibited sharp declines in late November when schools closed for U.S. Thanksgiving holiday and again in late December for Christmas vacation. When classes resumed in early January, *learning* web searches increased substantially. Web queries for *learning* gradually declined to the end of May/early June when the academic year ended (Figure 1). Lower levels of *learning*-seeking during the summer months are consistent with summer learning losses (Cooper et al. 1996). During the school week, daily *learning* queries were greatest on Mondays through Fridays when classes were typically in session, and searches were lowest on Saturdays and Sundays when classes did not meet (Figure 2). In the state of Rhode Island, *learning* information-seeking was markedly greater on days when students were in attendance at school (Figure 3). The Google query for *learning* was strongly correlated with web searches for traditional school subject and content areas, such as *math* ($R^2=0.7681$, Figure 4), *algebra* ($R^2=0.8127$, Figure 5), and *quadratic* ($R^2=0.8684$, Figure 6). Taken together, these findings suggest that *learning* query data is a reasonable proxy for monitoring seasonal learning.

The usefulness of *learning* search engine data as an indirect indicator of learning was evaluated by comparing it against Google Trends data for the term *motivation*. Google searches for *motivation* result in a variety of related sources, including inspirational quotes, stories that enhance self-efficacy, explanations of motivation, and motivation-promoting activities. Interventions that incorporate such elements have been reported to raise personal levels of motivation and influence behavior (Fisher et al. 1996; Stipek et al., 1998; Czuchry and Dansereau 2005; Hudley et al. 2007; Guilloteaux and Dörnyei 2008; Schelling et al. 2009). According to the Unified Learning Model (Shell et al. 2010) and other learning theories (Weiler 2005; Brooks and Shell 2006), motivation is the process that drives learning, and the two are intimately related. As predicted by these learning frameworks, *motivation*-seeking behavior exhibits similar temporal patterns as observed with *learning* (Figure 1 and Figure 7), and Google searches for the two terms are positively correlated ($R^2=0.6881$, Figure 8). In contrast, little correlation exists between *motivation* queries and searches for *exercise* ($R^2=0.1505$, Figure 9), *workout* ($R^2=0.0214$, Figure 10), *fitness* ($R^2=0.0882$, Figure 11), or *diet* ($R^2=0.0213$, Figure 12). The seasonal pattern of *motivation*-seeking and strong correlation with *learning* searches, suggest the same search engine user population is seeking *learning* and *motivation*.

The seasonal pattern of *learning* search activity suggests that this information-seeking is conducted predominantly by students and their families. A linguistic study of Twitter messages has also shown that community membership can be accurately predicted from online word usage (Bryden et al. 2013). Furthermore, research on search engine query data has shown that web searches for *flu* or *chicken pox* accurately mirror the number of flu and chicken pox patients (Ginsberg et al. 2009; Bakker et al. 2016). The 18.5% increase in *learning* queries over the 5-year study period, however, does not reflect the total number of students, as U.S. elementary and secondary school enrollment changed by only 0.2% and postsecondary enrollment rose by only 2.7% over the same time period (US DOE 2016). Instead, the increased volume of *learning* queries most likely reflects greater internet usage per student or family (Pew Research Center 2014) as well as the growth in online classes (Babson College 2015). The search engine trends reported in this paper are of aggregated data, so the correlations describe large populations rather than individual behavior. The aggregated Google data does not contain personal information on its searchers. Consequently, the identities of the search engine users, the reasons for searching, websites visited, and the level of internet accessibility are unknown.

Using less inclusive search terms, such as subject-specific educational topics, in future Google Trends studies may help to reveal some of the demographic characteristics of the *learning*-seeking population. Google search logs provide an abundant source of national and global data that can be used to complement traditional measures of learning. Internet search data has the added benefit of tracking information-seeking outside of traditional school settings, such as weekends, holidays, and summer, as well as beyond postsecondary education. Using web query data is a cost-effective, unobtrusive means of monitoring online learning, especially for school districts that lack testing resources but have sufficient internet capabilities.

Conclusions

Five-years of Google search engine queries for the term *learning* were studied for the entire United States population of internet users. The correspondence of *learning* web searches with the academic school calendar, association with student classroom attendance, and positive correlation with school subject and content areas all support the hypothesis that *learning* search engine queries can be used as an indirect measure of learning. Further support for the hypothesis was observed by the strong correlation between *learning*- and *motivation*-seeking behavior, as predicted by the Unified Learning Model and other learning frameworks.

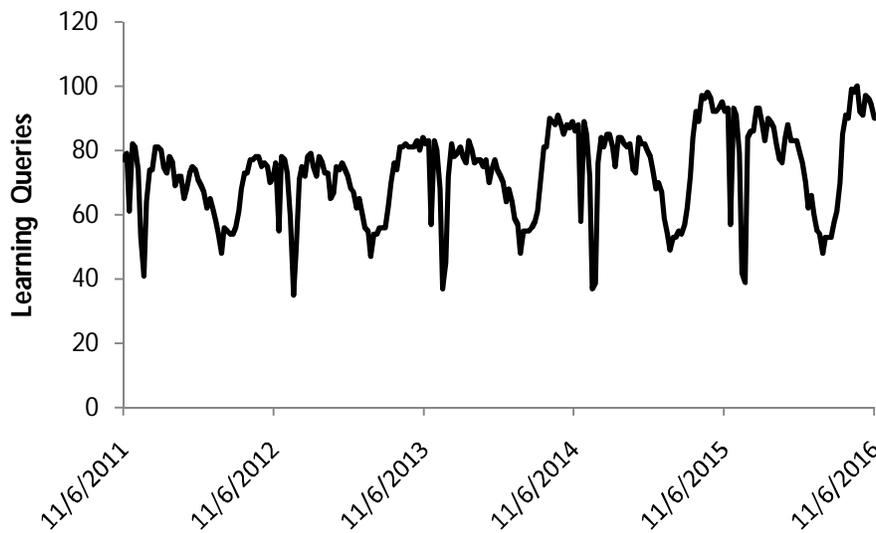


Figure 1. 5-year time course of Google *learning* queries for the United States population.

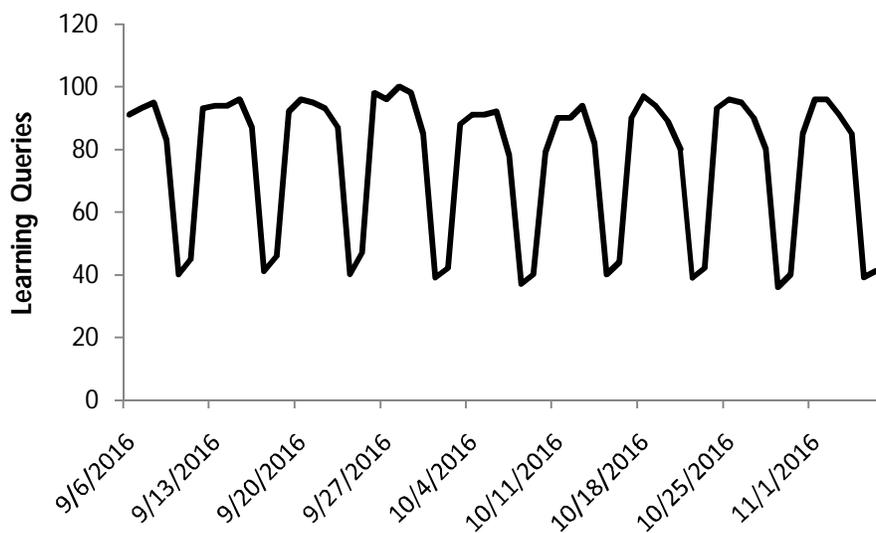


Figure 2. 60-day time series of Google *learning* queries for the United States population.

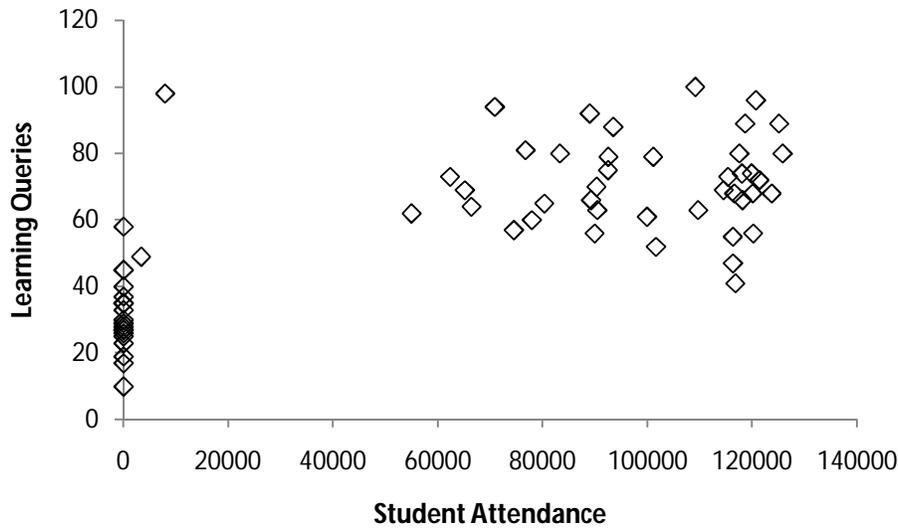


Figure 3. Relationship between Google search engine queries for *learning* and student attendance in K-12 classrooms in the state of Rhode Island.

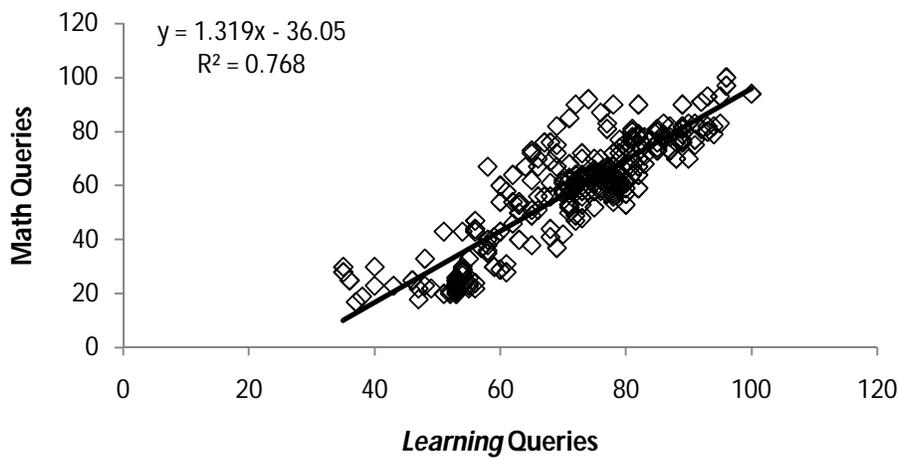


Figure 4. Association between Google search engine queries for *math* and searches for *learning*.

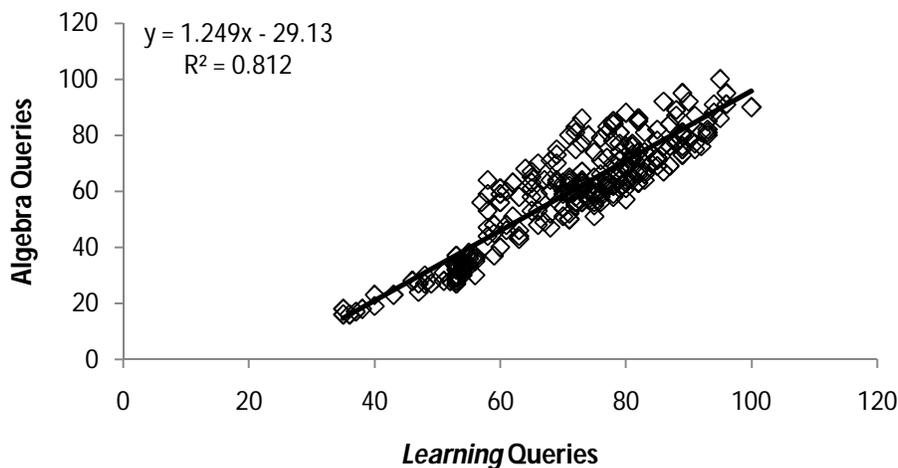


Figure 5. Google searches for *algebra* relative to searches for *learning*.

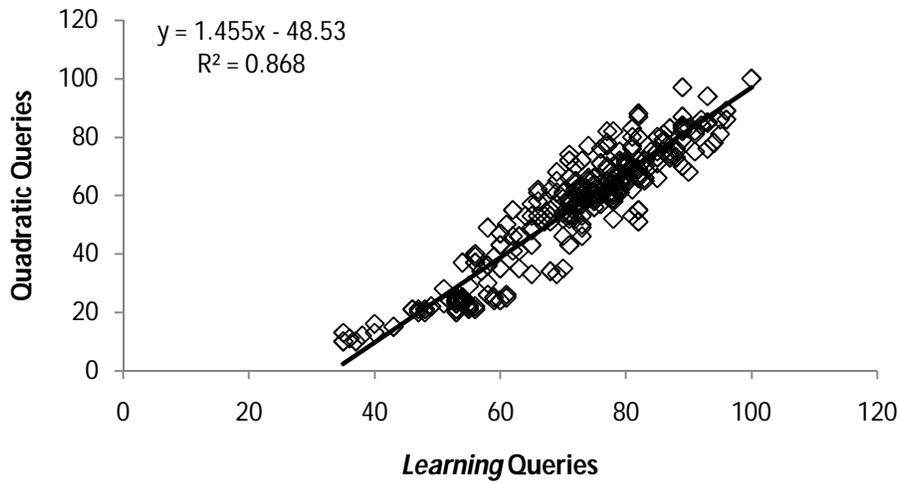


Figure 6. Relationship between *quadratic* web searches and *learning* searches.

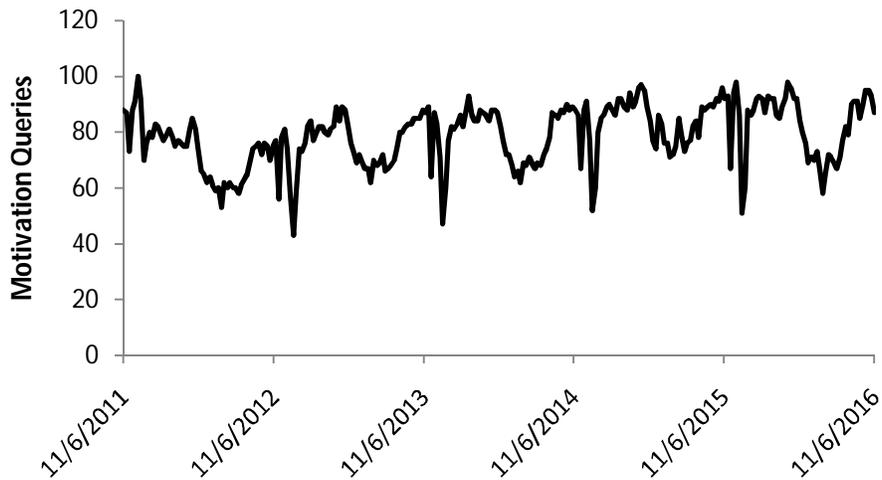


Figure 7. 5-year time course of Google queries for *motivation* in the United States.

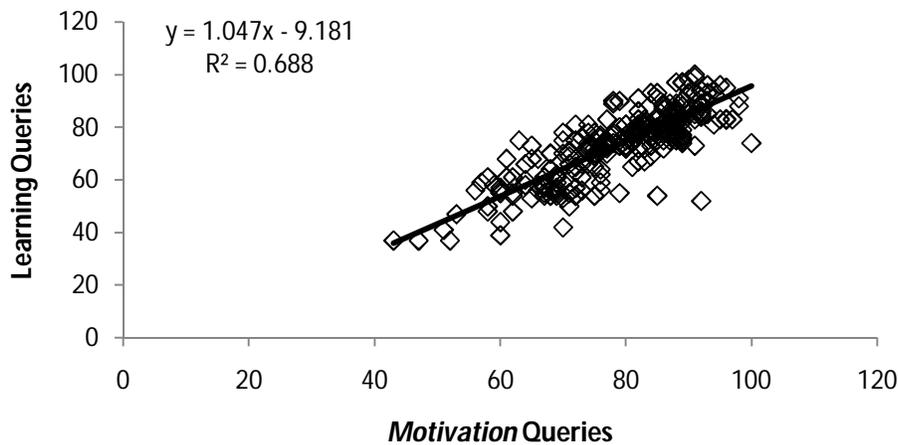


Figure 8. Relationship between Google search engine queries for *learning* and queries for *motivation*.

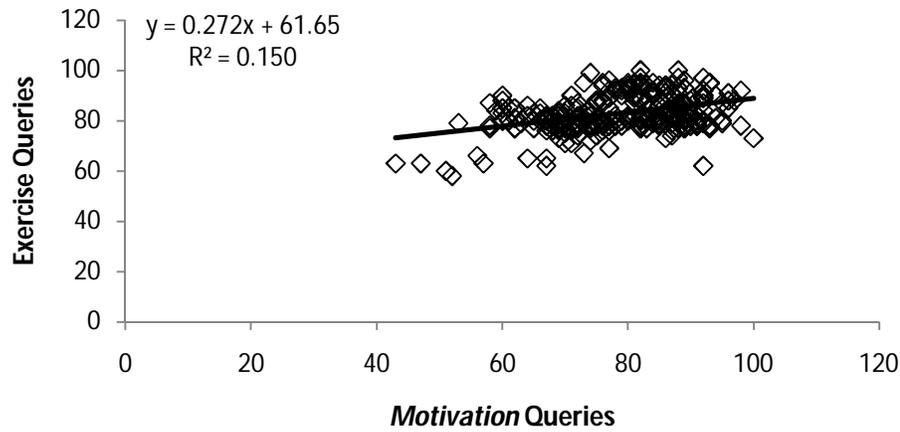


Figure 9. Google searches for *exercise* as a function of searches for *motivation*.

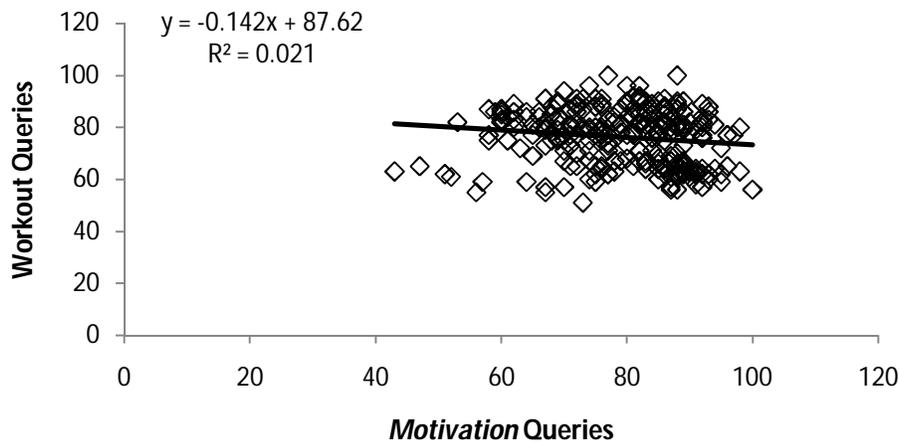


Figure 10. Google search engine queries for *workout* and *motivation*.

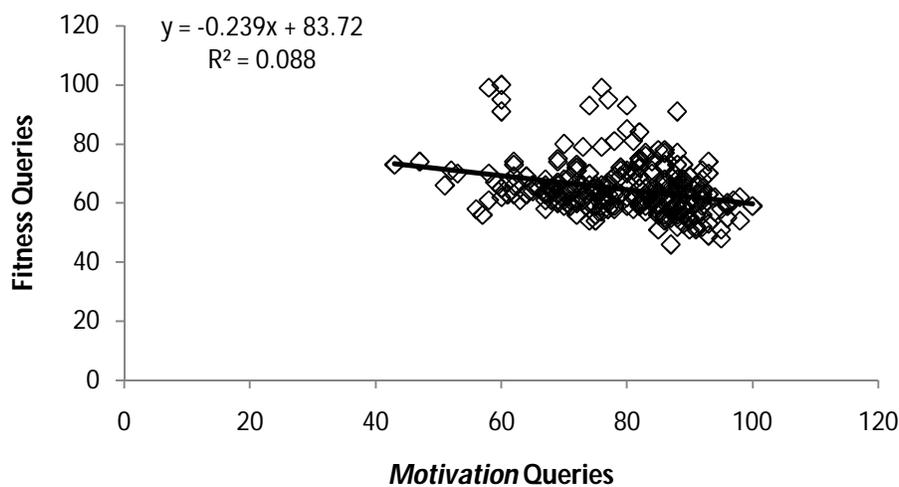


Figure 11. Google searches for *fitness* and *motivation*.

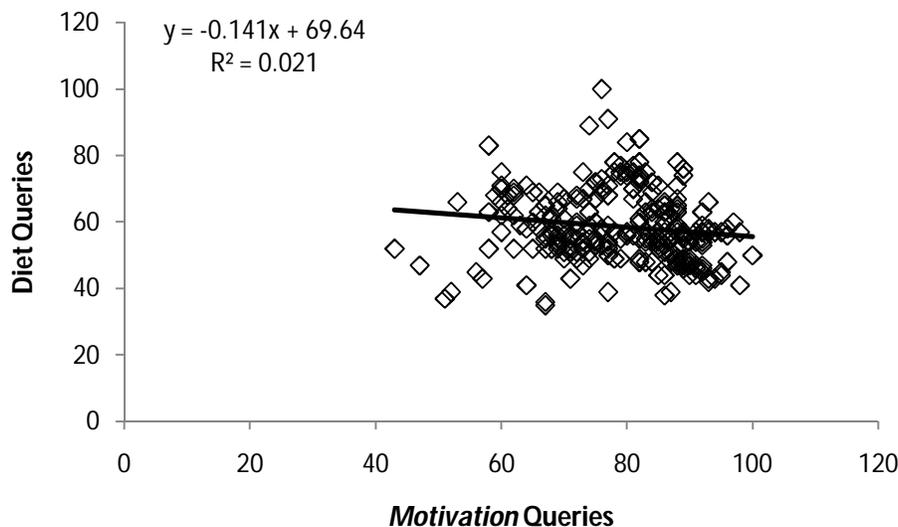


Figure 12. Diet and motivation Google queries.

References

- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. *Education Policy Analysis Archives*, 10, 18.
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers*. San Francisco: Jossey-Bass.
- Babson College (2015). Babson study: distance education growth continues. [Online] Available: <http://www.babson.edu/news-events/babson-news/Pages/2016-babson-releases-2015-survey-of-online-learning.aspx> (August 6, 2016)
- Bakker, K. M., Martinez-Bakker, M. E., Helm, B., & Stevenson, T. J. (2016). Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. *Proceedings of the National Academy of Sciences*, 113(24), 6689-6694.
- Brooks, D. W., & Shell, D. F. (2006). Working memory, motivation, and teacher-initiated learning. *Journal of Science Education and Technology*, 15(1), 17-30.
- Bryden, J., Funk, S., & Jansen, V. A. (2013). Word usage mirrors community structure in the online social network Twitter. *EPJ Data Science*, 2(1), 1.
- Cheng, X., Dale, C., & Liu, J. (2008, June). Statistics and social network of YouTube videos. In *Quality of Service, 2008. IWQoS 2008. 16th International Workshop on* (pp. 229-238). IEEE.
- Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(s1), 2-9.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Great house, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3), 227-268.
- Czuchry, M., & Dansereau, D. F. (2005). Using motivational activities to facilitate treatment involvement and reduce risk. *Journal of Psychoactive Drugs*, 37(1), 7-13.
- Desai, R., Hall, A. J., Lopman, B. A., Shimshoni, Y., Rennick, M., Efron, N., Matias, Y., Patel, M. M., & Parashar, U. D. (2012a). Nor virus disease surveillance using Google internet query share data. *Clinical Infectious Diseases*, 55(8), e75-e78.
- Desai, R., Lopman, B. A., Shimshoni, Y., Harris, J. P., Patel, M. M., & Parashar, U. D. (2012b). Use of internet search data to monitor impact of rotavirus vaccination in the United States. *Clinical Infectious Diseases*, 54(9), e115-e118.
- Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11), 87-92.

- Falvo, D. A., & Johnson, B. F. (2007). The use of learning management systems in the United States. *Tech Trends*, 51(2), 40-45.
- Fisher, J. D., Fisher, W. A., Misovich, S. J., Kimble, D. L., & Malloy, T. E. (1996). Changing AIDS risk behavior: effects of an intervention emphasizing AIDS risk reduction information, motivation, and behavioral skills in a college student population. *Health Psychology*, 15(2), 114.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012-1014.
- Google. (2016). Google trends. [Online] Available: <https://www.google.com/trends/> (November 6 2016)
- Guilloteaux, M. J., & Dörnyei, Z. (2008). Motivating language learners: A classroom-oriented investigation of the effects of motivational strategies on student motivation. *TESOL quarterly*, 55-77.
- Hudley, C., Graham, S., & Taylor, A. (2007). Reducing aggressive behavior and increasing motivation in school: The evolution of an intervention to strengthen school adjustment. *Educational Psychologist*, 42(4), 251-260.
- Hulth, A., Rydevik, G., & Linde, A. (2009). Web queries as a source for syndromic surveillance. *PloS one*, 4(2), e4378.
- Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, 42(1), 248-263.
- Jones, E. (2011). Google Books as a general research collection. *Library Resources & Technical Services*, 54(2), 77-89.
- Kuh, G. D. (2009). The national survey of student engagement: Conceptual and empirical foundations. *New directions for institutional research*, 2009(141), 5-20.
- Pew Research Center. (2014). Internet use over time. [Online] Available: <http://www.pewinternet.org/data-trend/internet-use/internet-use-over-time/> (November 6 2016)
- Polgreen, P. M., Chen, Y., Pennock, D. M., Nelson, F. D., & Weinstein, R. A. (2008). Using internet searches for influenza surveillance. *Clinical infectious diseases*, 47(11), 1443-1448.
- Rhode Island Department of Education. (2016) Daily attendance data. [Online] Available: <http://www.ride.ri.gov/informationaccountability/rieducationdata/dailyattendancedata.aspx>(August 6, 2016)
- Schelling, S., Munsch, S., Meyer, A. H., Newark, P., Biedert, E., & Margraf, J. (2009). Increasing the motivation for physical activity in obese patients. *International Journal of Eating Disorders*, 42(2), 130-138.
- Shell, D. F., Brooks, D. W., Trainin, G., Wilson, K. M., Kauffman, D. F., & Herr, L. M. (2010). *The Unified Learning Model*. Springer Netherlands.
- Shimshoni, Y., Efron, N., & Matias, Y. (2009). On the predictability of search trends. *Google Research Blog*.
- Stipek, D., Givvin, K. B., Salmon, J. M., & Mac Gyvers, V. L. (1998). Can a teacher intervention improve classroom practices and student motivation in mathematics? *The Journal of Experimental Education*, 66(4), 319-337.
- Tene, O. (2007). What google knows: Privacy and internet search engines. *Utah Law Review*.
- U.S. Department of Education, National Center for Education Statistics. (2016). *Digest of Education Statistics, 2014 (NCES 2016-006)*.
- Weiler, A. (2005). Information-seeking behavior in generation Y students: Motivation, critical thinking, and learning theory. *The Journal of Academic Librarianship*, 31(1), 46-53.