The size of an observer’s pupil varies with factors such as ambient luminance, arousal, cognitive effort, and working memory load. Since it is difficult to measure working memory load during an ongoing task, the ability to estimate it via pupil size recording could benefit many fields of research. One important paradigm for such application is visual search, whose utilization of working memory has been the subject of long-standing debates. Porter, Troschianko, and Gilchrist (QJEP 2007) found that pupil size statistically increased over the course of the search, and they attributed this finding to accumulating working memory load. However, other factors, e.g., arousal and effort, likely affected pupil size as well and added noise to the data and some uncertainty to the conclusions. In the present study, we interspersed a simple search task (find the "T" among "L"s) with intermittent blank screens showing only a central fixation marker, thought to induce a low, stable level of arousal and cognitive effort. Consequently, differences in minimum pupil size between successive fixation screens should mainly reflect changes in working memory load that occurred during the search interval between the screens. A within-subject analysis showed that this pupil size difference between the first two fixation screens was a significant predictor of RT in the same trial, with an inverse correlation of approximately $r = -0.4$, indicating that greater pupil size increase tended to be followed by shorter search time. Furthermore, the difference in average, minimum, or maximum pupil size between the first two search display presentations did not predict RT, $|r| < 0.07$. These results show that (1) working memory load increases during search, (2) this load increase is a major factor determining search efficiency, and (3) intermittent fixation screens greatly enhance pupil-based memory load estimation, even providing trial-by-trial insight into the utilization of working memory.